

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»**

**Факультет інформатики та обчислювальної техніки**  
(повна назва інституту/факультету)

**Кафедра автоматики та управління в технічних системах**  
(повна назва кафедри)

«На правах рукопису»  
УДК 81'322.4

«До захисту допущено»  
Завідувач кафедри  
\_\_\_\_\_  
(підпис) Ролік О. І.  
(ініціали, прізвище)  
“ \_\_\_\_ ” \_\_\_\_\_ 2018 р.

**Магістерська дисертація  
на здобуття ступеня магістра**

**зі спеціальності (спеціалізації) 126 «Інформаційні системи та технології»**  
(код і назва спеціальності)

---

на тему: Автоматизована система двомовного перекладу за допомогою маркерів

Виконав: студент VI курсу, групи ІА-72мп  
(шифр групи)

Кот Юрій Анатолійович  
(прізвище, ім'я, по батькові)

\_\_\_\_\_  
(підпис)

Науковий керівник д.т.н, професор, Теленик С. Ф.  
(посада, науковий ступінь, вчене звання, прізвище та ініціали)

\_\_\_\_\_  
(підпис)

Консультант \_\_\_\_\_  
(назва розділу) (науковий ступінь, вчене звання, прізвище, ініціали) (підпис)

Рецензент \_\_\_\_\_  
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали) (підпис)

Засвідчую, що у цій магістерській дисертації  
немає запозичень з праць інших авторів без  
відповідних посилань.

Студент \_\_\_\_\_  
(підпис)

Київ – 2018 року

**Національний технічний університет України  
«Київський політехнічний інститут  
імені Ігоря Сікорського»**

Факультет (інститут) \_\_\_\_\_ інформатики і обчислювальної техніки  
(повна назва)

Кафедра \_\_\_\_\_ автоматики та управління в технічних системах  
(повна назва)

Рівень вищої освіти – другий (магістерський)  
(код і назва)

Спеціальність \_\_\_\_\_ 126 «Інформаційні системи та технології»  
(код і назва)

**ЗАТВЕРДЖУЮ**

Завідувач кафедри

\_\_\_\_\_ **Ролік О. І.**  
(підпис) (ініціали, прізвище)

« \_\_\_\_\_ » \_\_\_\_\_ 2018 р.

**ЗАВДАННЯ  
на магістерську дисертацію студенту  
Коту Юрію Анатолійовичу**  
(прізвище, ім'я, по батькові)

1. Тема дисертації «Автоматизована система двомовного перекладу за  
допомогою маркерів» \_\_\_\_\_

науковий керівник дисертації Теленик Сергій Федорович, д.т.н., професор,  
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «29» жовтня 2018 р. № \_\_\_\_\_

2. Строк подання студентом дисертації \_\_\_\_\_ 4 грудня 2018 \_\_\_\_\_

3. Об'єкт дослідження процес машинного перекладу

4. Предмет дослідження автоматизована система німецько-українського  
перекладу з адаптованою статистичною моделлю перекладу

5. Перелік завдань, які потрібно розробити: проаналізувати існуючі  
системи та методи перекладу; спроектувати автоматизовану систему  
німецько-українського перекладу; сформулювати математичну модель  
системи перекладу за допомогою маркерів; провести експеримент на базі  
математичної моделі

6. Орієнтовний перелік ілюстративного матеріалу діаграма варіантів використання, діаграма розгортання, діаграма компонентів, проектування інтерфейсу користувача, діаграми баз даних, дві матриці відстеження

7. Орієнтовний перелік публікацій \_\_\_\_\_

8. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

9. Дата видачі завдання \_\_\_\_\_ 29 жовтня 2018 року \_\_\_\_\_

#### Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Строк виконання етапів магістерської дисертації	Примітка
1	Отримання завдання та узгодження вихідних даних	30.10.2018	
2	Вивчення об'єкту дослідження	10.11.2018	
3	Розробка математичної моделі	20.11.2018	
4	Розробка програмної моделі	25.11.2018	
5	Проведення експерименту	29.11.2018	
6	Оформлення документації	2.12.2018	
7	Подання роботи до попереднього захисту	4.12.2018	

Студент

\_\_\_\_\_  
(підпис)

Ю. А. Кот

(ініціали, прізвище)

Науковий керівник дисертації

\_\_\_\_\_  
(підпис)

С. Ф. Теленик

(ініціали, прізвище)

## РЕФЕРАТ

Магістерська дисертація освітньо-кваліфікаційного рівня “магістр” на тему «Автоматизована система двомовного перекладу на основі маркерів». 118 с., 6 рис., 66 табл., 9 додатків, 42 джерел.

Метою дисертації є розробка автоматизованої системи німецько-українського перекладу, створення математичної моделі перекладу за допомогою маркерів та проведення експерименту на математичній моделі.

В результаті виконання магістерської дисертації досягнуто мету, тобто розроблена системи, сформульована математична модель перекладу та проведено експеримент. Результати отримані внаслідок проведення експерименту на базі сформульованої моделі показують, що впровадження системи перекладу з використанням маркерів підвищує якість перекладу за загальноновживаною метрикою якості перекладу BLEU. В результаті порівняння якості перекладу проведеного між найпростішою статистичною моделлю, та створеною моделлю за рахунок тренування моделей на 3168 реченнях паралельного корпусу вдалося збільшити якість перекладу на 15%.

Математична модель може бути взята за основу подальших досліджень у збільшенні якості та зменшенні вибагливості до ресурсів систем статистичного перекладу та, теоретично, систем перекладу з використанням нейронних мереж.

Результати дисертації мають практичну цінність, що підтверджується відповідним актом їх впровадження на підприємстві ТОВ «АЙМАКС МЕНЕДЖМЕНТ УКРАЇНА», копію якого наведено у додатку А.

МАШИННИЙ ПЕРЕКЛАД, СТАТИСТИЧНИЙ ПЕРЕКЛАД, ГІБРИДНИЙ ПЕРЕКЛАД, ПАРАЛЕЛЬНИЙ КОРПУС ПЕРЕКЛАДІВ

## ABSTRACT

Master's dissertation of educational-qualifying level of "Master" on "Development of models, methods and means of automatic translation of scientific texts". 118 pgs., 6 fig., 66 tables, 9 applications, 42 sources.

The purpose of the dissertation is to develop an automated system of the German-Ukrainian translation, to create a mathematical translation model using markers and to conduct an experiment on mathematical model.

As a result of the master's thesis, the goal was achieved, which means, the system was developed, a mathematical translation model was described and an experiment was conducted. The results obtained as a result of the experiment on the basis of the formulated model show that the introduction of a translation system using markers increases the quality of translation measured by commonly used translation quality metrics BLEU. As a result of a comparison of the quality of the translation carried out between the simplest statistical model and the model trained on 3168 sentences of parallel corpora the quality translation increased by 15%.

A mathematical model can be taken as the basis for further research in increasing the quality and reducing the exactingness of statistical translation systems and, theoretically, the of neural network translation systems.

The results of the thesis have practical value, which is confirmed by the relevant act of their implementation at the company "АЙМАКС МЕНЕДЖМЕНТ УКРАЇНА" LLC. Scan of this document is provided in application A.

MACHINE TRASLATION, STATISTICAL TRANSLATION, HYBRID  
TRANSLATION, PARALLEL CORPORA

## ЗМІСТ

ВСТУП.....	9
1 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ .....	12
1.1 PROMT Translator.....	12
1.2 Google Translate .....	14
1.3 Microsoft Translator.....	20
1.4 Висновки .....	26
2 ПОНЯТТЯ МАШИННОГО ПЕРЕКЛАДУ .....	27
2.1 Вступ.....	27
2.2 Машинний переклад на базі правил.....	29
2.3 Статистичний машинний переклад .....	31
2.4 Машинний переклад на базі нейронних мереж .....	34
2.5 Гібридні системи перекладу.....	37
2.6 Висновки .....	39
3 ФОРМУВАННЯ ВИМОГ ДО СИСТЕМИ ТА ПРОЕКТУВАННЯ.....	43
3.1 Формування вимог до системи .....	43
3.2 Сценарії використання системи.....	47
3.3 Розгортання системи та бізнес вимоги .....	59
3.4 Компоненти системи.....	61
3.5 Обґрунтування вибору компонентів .....	62
3.6 Структура бази даних .....	63
3.6 Розроблення інтерфейсу користувача.....	66
3.7 Тестування системи .....	68
3.8 Висновки .....	72

4 МАТЕМАТИЧНА МОДЕЛЬ ПЕРЕКЛАДУ ТА ЕКСПЕРИМЕНТ .....	73
4.1 Поняття перекладу за допомогою маркерів .....	73
4.2 Приклад лінгвістичної бази даних .....	75
4.3 Виконання експерименту .....	78
4.4 Висновки .....	82
5 РОЗРОБКА СТАРТАП-ПРОЕКТУ .....	83
5.1 Опис ідеї стартап-проекту .....	83
5.2 Технологічний аудит проекту .....	89
5.3 Аналіз ринкових можливостей запуску проекту .....	91
5.4 Розроблення ринкової стратегії проекту .....	106
5.5 Розроблення маркетингової програми стартап-проекту .....	110
5.6 Висновки .....	115
ВИСНОВКИ .....	117
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ .....	119
ДОДАТОК А	
ДОДАТОК Б	
ДОДАТОК В	
ДОДАТОК Г	
ДОДАТОК Ґ	
ДОДАТОК Д	
ДОДАТОК Е	
ДОДАТОК Є	
ДОДАТОК Ж	

## СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ

МП – Машинний переклад

SMT – Statistical Machine Translation (Статистична система машинного перекладу)

RBMT – Rule-based Machine Translation (Система машинного перекладу заснована на правилах)

PBMT – Phrase-based Machine Translation (Система машинного перекладу заснована на фразях)

HMT – Hybrid Machine Translation (Гібридна система машинного перекладу)

СМП – Система машинного перекладу

ОС – Операційна система

CTF – Collaboration translation framework (Фреймворк спільного перекладу)



## ВСТУП

Швидше за все ми не скоро зможемо побачити перекладач, який має якість перекладу таку ж або краще ніж у професійного перекладача. Люди, які володіють мовами, здатні передавати зміст написаного своїми словами, не прив'язуючись до структури першоджерела. Машини у свою чергу переводять послівно або пофразово і навчити їх оперувати не словами, а образами — це все одно, що винайти штучний інтелект. Оперувати образами означає розуміти текст, що перекладається та інтерпретувати його. Хоча за більш ніж 70 років існування машинного перекладу ми вже пройшли досить великий шлях від статистичних методів до штучних нейронних мереж.

Нейронні мережі вміють читати речення і зліва направо, і справа наліво, по літерам транслітерувати власні імена і замість того, щоб запам'ятовувати безліч варіантів перекладу, оперують семантикою цілого тексту, розбиваючи його на сегменти, після чого аналізують і синтезують їх. Результат виходить якісний у порівнянні з іншими системами, і, в деяких випадках, система переводить навіть фразеологізми [1].

Але навіть нейронні мережі не можуть впоратися з усіма труднощами перекладу. Мова — це дуже гнучка система з необмеженим набором розмитих правил. І хоча нейронні мережі вже схоплюють семантичні і синтаксичні зв'язки в реченнях і навіть розпізнають акцент мовця, вони не вміють враховувати культурологічні, когнітивні, літературні та інші аспекти перекладу. Іншими словами, контекст може кардинальним чином порушити комунікацію, тому що:

- Комп'ютер не може зрозуміти культуру.
- Комп'ютер погано переводить ідіоми.
- Комп'ютер не здатний передати емоції пісень і віршів.
- Емоційно-забарвлені тексти, фразеологічні звороти, культурний підтекст також є важким для машини.

- Помилка у договорах, гарантійних листах, маркетингових матеріалах, медичних документах може коштувати комусь життя. В цьому випадку не можна цілком покладатись на машину.
- Рекламні гасла, будь-які художні тексти — це занадто важко, багатозначно і не формалізовано для машинного перекладу.

Машинний переклад в основному призначено для людей, які не знають мови, якою написаний текст, але яким потрібно в найзагальніших рисах зрозуміти зміст цього тексту. Також для перекладачів, яким необхідний «шаблон» для редагування. І для бізнесу, якому потрібно локалізувати свій контент різної глибини та складності.

Машинний переклад все одно доведеться редагувати людині, а для цього потрібно вміти помічати і виправляти помилки, зроблені машиною. Це окремий трудомісткий процес, який вимагає специфічного досвіду.

Навіть останні розробки машинного перекладу на базі нейронних мереж змогли підвищити BLEU оцінку якості перекладу тільки на 10% [2].

Як вже говорилося раніше, «універсальний» перекладач не завжди дає прийнятну якість і не може підтримувати специфічну термінологію. Щоб інтегрувати в свої процеси і застосовувати нейронні мережі для перекладу, потрібно виконати основні вимоги:

- Наявність достатніх обсягів паралельних текстів для того, щоб мати можливість навчати нейронну мережу. Часто у замовника їх просто мало або взагалі текстів з даної тематики не існує в природі. Вони можуть бути засекречені або знаходиться в стані не дуже придатному для автоматичної обробки. Для створення моделі потрібна база, де міститься мінімум 100 млн. слововживань, а щоб отримати переклад більш-менш прийнятної якості - 500 млн. слововживань. Далеко не кожна компанія володіє таким обсягом матеріалів.
- Наявність механізму або алгоритмів автоматичної оцінки якості одержуваного результату.
- Достатні обчислювальні потужності.

«Універсальний» нейронний перекладач найчастіше не підходить за якістю, а щоб розгорнути свою приватну нейронну мережу, здатну забезпечити прийнятну якість і швидкість роботи, потрібна «маленька хмара» що має трудомісткий процес налаштування та підтримки.

Ми приходимо до висновку, що для кожного, окремо взятого, завдання завдання з перекладу потрібно використовувати той перекладач, який максимально для цього підходить.

З урахуванням вищесказаного актуальність проблеми стає очевидною - у процесі перекладу залишилось багато невирішених проблем та задач у автоматизованих систем перекладу та методах перекладу.

## 1 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ

### 1.1 PROMT Translator

Це російська компанія, розробник систем машинного перекладу. Займається дослідженнями і розробками в області штучного інтелекту. Основні напрямки роботи – розробка рішень для машинного (автоматичного) перекладу і технологій для аналізу текстових неструктурованих даних російською та іноземними мовами.

Технології PROMT не раз отримували престижні міжнародні нагороди – так, PROMT Translation Server був визнаний кращим рішенням для фінансового сектора на саміті IDC Financial Insight, а PROMT Professional – «найбільш повним рішенням» і переможцем в категорії «Якість» по версії журналу PC & I в Іспанії. У період з 2013 до 2016 року переклади PROMT з англійської на російську отримували кращі оцінки експертів Асоціації комп'ютерної лінгвістики (ACL).

Компанія заснована в Санкт-Петербурзі в 1991 році колишніми співробітниками лабораторії інженерної лінгвістики ЛГПІ ім. А. І. Герцена. Керівник компанії з дня заснування – Світлана Соколова.

6 березня 1998 року PROMT запускає перший російський сервіс безкоштовного онлайн-перекладу Translate.ru. У 1997 році PROMT разом з французькою компанією Softissimo випускає лінійку продуктів для західного ринку під назвою Reverso. У 2008 році контрольний пакет акцій компанії був викуплений Renova Capital. Сума угоди не розголошується, але за даними «Ренова», загальний обсяг світового ринку послуг перекладу для компаній склав в 2007 році \$ 9-10 млрд, при цьому продажі PROMT склали 2% від цієї суми, тобто \$ 180-200 млн. Виходячи з цих даних, за оцінками одного зі спостерігачів ринку, вартість придбаного «Ренова» пакета склала не більше \$ 20-40 млн.

Спочатку в основі перекладу PROMT лежала технологія машинного перекладу, заснована на правилах (Rule-based machine translation). RBMT будується на основі лінгвістичного опису двох природних мов (двомовних словників та інших баз даних, що містять морфологічну, граматичну та семантичну інформацію), формальних

граматик і, власне, алгоритмів перекладу. Якість перекладу залежить від обсягів лінгвістичних баз даних (словників) і глибини опису природних мов, тобто, необхідний облік максимальної кількості особливостей граматичної структури. RBMT підхід був удосконалений в PROMT за рахунок використання семантичних мереж і отримав назву Аналітичний машинний переклад PROMT. Він характеризується високою здатністю системи до навчання на переклад термінології, високою швидкістю роботи і надійністю.

У продуктах PROMT також використовується статистичний машинний переклад і переклад на основі нейронних мереж. У першому випадку навчання відбувається на основі значних за обсягом корпусів паралельних текстів – і якість перекладу сильно залежить від якості і кількості даних, результат відрізняється більшою гладкістю, ніж переклади, виконані на основі аналітичного підходу

Нейронний підхід відрізняється високою якістю перекладу, близькому до перекладу, виконаному людиною, можливістю перекладу контенту, який відсутній в навчальних даних і можливістю навчання системи на переклад спеціальної термінології.

Основні можливості:

- Переклад окремих слів і текстів будь-якого обсягу.
- Переклад документів зі збереженням структури і форматування. Підтримуються всі популярні формати: doc(x), xls(x), ppt(x), rtf, html, xml, txt, ttx, pdf(в тому числі відскановані), odt, ods, jpeg, png, tiff.
- Переклад сайтів цілком, зі збереженням структури і гіперпосилань.
- Переклад виділених фрагментів тексту і окремих слів в сторонніх додатках і браузерах (PROMT Агент).
- Використання, редагування і створення спеціалізованих словників і профілів перекладу.
- Підключення баз Translation Memory (пам'ять перекладу).
- Інтеграція в офісні додатки, веб-браузери, корпоративні портали і сайти.
- Автоматичне розпізнавання мови, тематики тексту.

- Забезпечення конфіденційності перекладу даних для клієнтів, яким це важливо за рахунок роботи в офлайн-режимі.
- Підтримка MacOS, Windows, Linux, Android, iOS.
- API і SDK для інтеграції функцій перекладу в програмні комплекси та веб-сторінки.
- API і SDK для інтеграції функцій аналізу текстів (виділення сутностей, фактів, зв'язків між ними, визначення та аналіз смислової структури документа) в інформаційно-аналітичні системи.

## 1.2 Google Translate

В 2011 році переклад в багатьох мовних парах йшов через проміжний переклад на англійську з ефектом «зламаного телефону».

При перекладі в зазначених напрямках англійська мова виступала «посередником»: текст спочатку перекладався англійською і лише потім на вибрану мову перекладу. На мову перекладу перекладався спотворений англійський варіант, з неминучими при машинному перекладі неточними фрагментами. Так на «первинну деформацію» накладався другий шар. В результаті один і той же текст отримував більше спотворення при перекладі на німецьку, французьку та ін. мови, ніж на англійську.

Коли порядок слів в англійському тексті порушений, форми слів і їхня послідовність не є надійним індикатором ролі цих слів у реченні [при подальшому перекладі на цільові мови].

Статистичний переклад оптимальний між спорідненими мовами. Переклад Google з російської на українську і навпаки найбільш «інтуїтивно» вірний з усіх розглянутих, найближчий до готового продукту, в ньому менше спотворення сенсу або правил мови, менше дослівного перекладу. Правильна «конвертація» граматичних конструкцій однієї мови засобами іншої — це межа для статистичного машинного перекладу. Це обмеження не скасовується навіть на базі споріднених мов

і створює тим більше «шуму» в інтерпретації, чим менше граматичне схожість мов в парі.

Англійська — «центральна» мова в Google Translate. Переклад з англійської та на англійську в Google Translate — прямий, без посередництва іншої мови. Це забезпечувало непогану якість, при якій особливо помітні гідності роботи сервісу: часто правильний переклад імен та назв, термінів, фразеологізмів, використання живих мовних зворотів на відміну від дослівного перекладу, нерідко правильний вибір лексичного значення в залежності від контексту. В 2016 році Google прибирає повтори з перекладу; краще вибудоване, зв'язне речення, іноді більш вдалий вибір слів; іноді ставався відкат на менш вдалий переклад («камінь інструмент» замість «кам'яне знаряддя» в 2011 році); іноді менш вдала інтерпретація ролі члена речення — іноді більше. Таким чином, машинні переклади в 2016 році були місцями краще, місцями гірше, ніж переклади 2011 року, але в цілому рівень той самий [3].

В 2017 році англійська як мова-посередник зберігала свою роль, але потрохи здавала позиції. З'явилося більше варіацій, відхилень від англійського проміжного перекладу. Часто ці експерименти були невдалими, тобто якби переклад на мову перекладу, як і раніше наосліп йшов за англійським, результат був би кращий. Однак паралельно покращилося «володіння» граматикою цільової мови. Структура речень і вибір лексики на цільових мовах як і раніше багато в чому визначався англійським перекладом.

Є тенденція до генерації тексту цільовою мовою відповідно до законів його граматики. Кореляції між перекладами одного тексту на різні мови менше, ніж раніше. Сервіс не переводив дослівно, результат ставав більш вільним: адекватна перефразування, перегруповання слів, перестановка слів з початку в кінець речення, якщо того вимагають правила мови.

На відміну від попереднього рівня (phrase-based translation — одноразове знаходження відповідностей окремих слів і фраз), нейронний перекладач в якійсь мірі трансформував пропозиції, аналізував їх як єдине ціле і встановлював відповідності «від краю до краю» в кілька стадій (end- to-end mapping — наскрізне перетворення, повного циклу, безперервна трансформація різноманіття даних зі входу на вихід).

Головне досягнення в результатах переведення 2017 року — більш тверде, впевнене розпізнавання структури речення і передача граматичних значень на цільових мовах. В англійській мові закінчення не грають таку важливу роль в передачі граматичних значень, як в українській, німецькій, польській та російській мовах. Проте, при «прогоні» через нейромережу граматичні зв'язки стали «губитися» рідше, ніж при статистичному перекладі. Також стали розпізнаватися рідковживані багатокореневі слова: перекладач добре справлявся з членуванням не тільки речення, але і слова.

За період у 2017-2018 роках якість машинного перекладу значно зросла. У 2011 - 2016 роках в перекладах складних фраз англійською була тільки видимість зв'язності: перекладені слова і словосполучення нанизувалися в ланцюжок в злегка скоректованому порядку, але «глибинного розуміння» структури не було, і часом переклад виглядав адекватним тільки тому, що в англійській часто не потрібні закінчення, а відсутність службових слів в деяких стилях допустима. Зате це «нерозуміння» завжди виявлялося в подальших перекладах на цільові мови. У перекладах за 2017 рік структура англійського речення вивірена краще — і краще інтерпретується на інші мови. Якість на цих мовах підвищилося пропорційно: трохи нижче англійської, але набагато вище попередніх варіантів.

У 2011 сервіс іноді сприймав шматки складного речення як ізольовані і просто нанизував їх переклад одне за одним у ланцюжок. У 2017 році, цю проблему вирішено, також краще ізолюються дійсно сторонні шматки, щоб вони не створювали «шуму». Це вкраплення слів іншою мовою та помилки. Це наближає машину до рівня людини: якщо ми не почуємо кілька слів у реченні, як правило, це не заважатиме вловити загальний зміст.

Нейронний перекладач не оперує смислами. Статистичний перекладач добре працював з розпізнаванням термінів, імен, фраз, часто вдало вибирав значення слів в контексті речення. Проблеми починалися, коли не виходило правильно інтерпретувати взаємозв'язок між словами, їхню граматичну роль. У перекладах 2017 року помітне значне поліпшення в цьому напрямку, тобто нерозпізнаних місць стало



менше. Нейронний перекладач було прив'язано до структури речення. Він почав непогано трансформувати мовні конструкції з мови на мову.

Функції Google Translator наведені у таблиці 1.1.

Таблиця 1.1 - Функції Google Translate

Функція	Опис
Письмові переклади слів	Функція, яка перекладає письмові слова чи текст на іноземну мову [4].
Переклад веб-сайту	Функція, яка перетворює цілу веб-сторінку на вибрані мови
Переклад документів	Функція, яка перекладає документ, завантажений користувачами до вибраних мов. Документи мають бути такими: .doc, .docx, .odf, .pdf, .ppt, .pptx, .ps, .rtf, .txt, .xls, .xlsx [5]
Переклад мовлення	Функція, яка миттєво перекладає розмовну мову на обрану іноземну мову [6].
Переклад мобільного додатка	У 2018 році Google Перекладач представив свою нову функцію під назвою "Натисни, щоб перекласти", яка зробила миттєвий переклад доступним всередині будь-яких додатків без виходу або перемикання його. [7]
Переклад зображення	Функція, яка ідентифікує текст у знімку, зроблену користувачами, і перетворює текст на екрані миттєво. [8]
Рукописний переклад	Функція, яка перекладає мову, написана на екрані телефону або

Функція	Опис
	накладається на віртуальну клавіатуру без підтримки клавіатури [9].

Додаткові можливості описані у таблиці 1.2.

Таблиця 1.2 - Додаткові можливості Google Translator

Назва можливості	Опис можливості
Інтеграція браузера	Перекладач Google доступний у деяких веб-переглядачах як додаткове завантажуване розширення, яке може запускати движок перекладу. [21] У лютому 2010 року перекладач Google був інтегрований в браузер Google Chrome за замовчуванням для додаткового автоматичного перекладу веб-сторінок [10].
Мобільні додатки	<p>Програма Google Translate для Android і iOS підтримує більше 100 мов та може перекладати 37 мов через фото, 32 - голосом у режимі "бесід" та 27 - у реальному часі у режимі "розширеної реальності". [11]</p> <p>Додаток для Android був випущений в січні 2010 року, а для iOS - 8 лютого 2011 року.</p> <p>Версія Android в січні 2011 року експериментувала з "Режимом розмови", що дозволило користувачам швидко спілкуватися з близькою людиною іншою мовою. Спочатку вона обмежувалася англійською та іспанською мовами, ця функція отримала підтримку для 12 нових мов, які ще перебувають у тестуванні, наступного жовтня.</p>

Назва можливості	Опис можливості
	<p>У січні 2015 року додатки отримали можливість перекладати фізичні знаки в реальному часі за допомогою камери пристрою, в результаті придбання компанією програми Word Lens. [12] Початковий запуск січня підтримував лише сім мов, але липнєве оновлення додало підтримку для 20 нових мов, а також підвищило швидкість перекладу режиму розмови. [13]</p>
API	<p>У травні 2011 року компанія Google оголосила, що API для Google Translate для розробників програмного забезпечення було застарілим і припинило функціонувати. Сторінка API для перекладу пояснила причину "суттєвим економічним навантаженням, спричинену великим зловживанням", з датою закінчення, встановленої на 1 грудня 2011 року. У відповідь на громадський тиск, Google оголосила в червні 2011 року, що API продовжуватиме бути доступним як платна послуга.</p> <p>Оскільки API використовувався на безлічі сторонніх веб-сайтів та додатків, оригінальне рішення знецінити це призвело до того, що деякі розробники критикували компанію Google і сумнівалися в можливості використання API Google у своїх продуктах. [14]</p>
Google Асистент	<p>Перекладач Google також пропонує переклади для Google Assistant та пристрої, які використовує Google Assistant, наприклад Google Home і Google Pixel Buds.</p>

Перекладач Google може перекладати з різних мов на різні форми текст та медіа, включаючи текст, мову, зображення, сайти або відео в реальному часі. [15] Він підтримує понад 100 мов на різних рівнях, а станом на травень 2017 року обслуговує понад 500 мільйонів людей щодня. [16] Для деяких мов програма Google Translate може вимовляти перекладений текст, виділити відповідні слова та фрази в тексті джерела та цілі та служити простою словником для введення одного слова. Якщо вибрано "Визначити мову", текст на невідомому мовою може бути автоматично ідентифікований. Якщо користувач вводить URL-адресу у вихідному тексті, Google Translate створює гіперпосилання на машинний переклад веб-сайту. [17] Користувачі можуть зберігати переклади в "розмовнику" для подальшого використання. Для деяких мов текст можна вводити за допомогою екранної клавіатури, за допомогою розпізнавання рукописного тексту або розпізнавання мови. [18]

### 1.3 Microsoft Translator

Перша версія системи машинного перекладу Microsoft була розроблена в період з 1999 по 2000 рр. в межах Microsoft Research. Ця система була заснована на семантичних структурах предикатів-аргументів, відомих як логічні форми (LF), і була відвернута від граматичної корекції, розробленої для Microsoft Word. Ця система, зрештою, використовувалася для перекладу всієї бази знань Microsoft на іспанську, французьку, німецьку та японську. Підхід Microsoft до машинного перекладу, як і більшість сучасних систем машинного перекладу, є "керованим даними", замість того, щоб покладатися на написання явних правил для перекладу природної мови, алгоритми навчаються розуміти та інтерпретувати перекладені паралельні тексти, дозволяючи їм автоматично навчитися перекладати новий текст мови. Досвід Microsoft з LF-системою призвів безпосередньо до деревоподібної системи перекладу, яка спростила LF до дерев залежностей, і в кінцевому підсумку до моделі шаблону замовлень, що призвело до значного покращення швидкості та включення нових цільових мов.

Переклад з використанням споживачів, відомий як Bing Translator (раніше відомий як перекладач Windows Live), був запущений у 2007 році та надає безкоштовний переклад тексту та веб-сторінок в Інтернеті. Текст перекладається прямо на веб-сторінці Bing Translator, а веб-сайти перекладаються за допомогою інструментів двомовного перегляду.

У 2011 році служба була розширена, щоб включати численні продукти Microsoft Translator через хмарне API, яке підтримує продукти, доступні як для споживачів, так і для корпоративних користувачів. Додаткова можливість мовлення була введена в березні 2016 року.

У травні 2018 року було оновлено API. Ця нова версія запропонувала нейронний машинний переклад як метод перекладу за замовчуванням. На додаток до перекладу, нова версія має транслітерацію та двомовний словник, щоб шукати слова, знаходити альтернативні переклади та переглядати приклади у реченнях.

Microsoft Translator використовує машинний переклад для створення миттєвих перекладів з однієї природної мови на іншу. Його система заснована на чотирьох різних областях комп'ютерного навчання досліджень.

Таблиця 1.3 – Моделі перекладу Microsoft Translator

Тип навчання	Вплив на Переклад
Нейронна мережа	Нейронна мережа намагається перекладати між мовами так само як працює мозок людини. На високому рівні, нейронні мережа переклад працює в два етапи. На першому етапі моделюється слово, що треба перекласти на основі контексту цього слова (і його можливий перекладів) в межах в речення. По-друге, нейронна мережа перекладає модель цього слова (не слово саме по собі а модель, побудовану нейронною мережею), в межах контексту речення, в іншу мову. [19] Переклад на базі нейронних мереж за

Тип навчання	Вплив на Переклад
	замовчуванням використовується в Microsoft API перекладача .
SMT на базі синтаксису	Переклад на основі синтаксису полягає в перекладі синтаксичних одиниць більше ніж переклад слів бо рядків. Microsoft використовував Синтаксичний SMT щоб перекласти тексти пов'язані з комп'ютерними текстами з англійської на інші мови. Постійні дослідження в цій сфері привели до великих поліпшень у відображенні слів та у перестановці слів.
SMT на основі фраз	В SMT на основі фраз, машина вчиться знаходити відповідності перекладу у корпусі паралельних текстів співставлення речення до речення між двома мовами, без потреби у лінгвістичному поясненні оброблюваних даних. Це дозволяє створити якісні системи перекладу за менший час ніж інші системи.
Вирівнювання слів	SMT системи покладаються на існуючі дані про переклад щоб навчитись як автоматично перекладати з однієї мови на іншу. Щоб навчити такі системи ідентифікація співставлення слів вихідної на цільової мови це важливий момент. Microsoft зробив велику роботу в дискримінативному та генеративному підходах для того щоб співставити слова. Це призвело до швидших алгоритмів та до збільшення якості перекладів.
Моделювання мови	Мова моделювання використовує n-грам моделі для того, щоб побудувати зрозумілий переклад на цільову мову. Це забезпечує те, що на виході

Тип навчання	Вплив на Переклад
	перекладу можемо мати читабельний та природній текст без помітного машинного акценту.

Якість результатів машинного перекладу Microsoft Translator оцінюється за допомогою методу, що називається оцінка BLEU [20].

BLEU (двомовна оцінка недосконалості) – це алгоритм оцінки якості тексту, який було автоматично перекладено з однієї природної мови на іншу. Якість вважається відповідністю між продуктивністю машини та продуктивністю людини. BLEU була однією з перших показників для досягнення високої кореляції з людськими судженнями про якість і залишається однією з найпопулярніших автоматизованих та недорогих показників.

Оскільки машинний переклад заснований на статистичних алгоритмах, а не на перекладачах, автоматичні переклади, які він створює, не завжди цілком точні. Microsoft Translator запровадив різні можливості зворотного зв'язку, такі як CTF, у свої продукти, щоб дозволити користувачам запропонувати альтернативні переклади. Потім ці альтернативні переклади інтегруються в алгоритми Microsoft Translator для вдосконалення майбутніх перекладів.

У листопаді 2016 року Microsoft Translator представив переклад з використанням нейронних мереж для дев'яти мов з найбільшим використанням у світі, включаючи японську мову. Нейромережі забезпечують кращий переклад, ніж галузеве стандартне статистичне навчання машин [21].

Microsoft Translator містить у собі певні ключові сервіси. Microsoft Translator – це хмарне API, інтегроване в численні продукти та служби Microsoft. Перекладач API може бути використаний самостійно, і може бути налаштований для використання в середовищі попереднього видання чи після публікації. API, доступний через підписку, безкоштовний для нижчих обсягів перекладу та стягується відповідно до багаторівневої системи оплати за обсяги, що перевищує два мільйони символів на

місяць. Решта основних продуктів доступні безкоштовно. Перелік основних продуктів від Microsoft Translator приведений у таблиці 1.4.

Таблиця 1.4 – Основні продукти Microsoft Translator

Назва функції	Опис
Хмарний перекладач Microsoft Translator	<p>Microsoft Translator – це служба автоматичного перекладу в хмарі, яка може використовуватися для створення додатків, веб-сайтів та інструментів, що потребують багатомовної підтримки.</p> <p>Текстовий переклад: Microsoft Translator Text API може використовуватися для перекладу тексту на будь-які мови, підтримувані службою.</p> <p>Мовний переклад: Microsoft Translator інтегрований в службу Microsoft Speech, який є кінцевим API на базі REST сервісів, який може бути використаний для створення додатків, інструментів або будь-якого рішення, що потребує мовленнєвого перекладу з багатьма мовами. Переклад мовлення до мови доступний для будь-якої розмовної мови або з будь-якої розмовної мови, а переклад мовлення у текстовий вигляд перекладу доступний з розмовних мов у будь-яку підтримувану мову.</p>
Користувацький перекладач	<p>Користувацький перекладач – це особливість служб Microsoft Translator, яка дозволяє підприємствам, розробникам програм та постачальникам мовних послуг створювати системи нейронних перекладів, які розуміють термінологію, яка використовується у власному бізнесі та галузі. Потім користувацькі переклади можна додати в</p>



Назва функції	Опис
	існуючі програми, робочі процеси та веб-сайти, використовуючи звичайний виклик до Microsoft Translator REST API. Користувальницький перекладач може використовуватися при перекладі тексту за допомогою Microsoft Translator Text API та при перекладі мови за допомогою служб Microsoft Speech [22].
Live feature	Ця функція являє собою особистий універсальний перекладач, який дозволяє людям в обсязі до 500 чоловік, з багатьма видами пристроїв, серед багатьох мов (багатомовних), мати особисто перекладені бесіди [23]. Ця функція наразі є безкоштовною і доступна в додатках Microsoft Translator (Android, iOS або Windows) та браузері за посиланням: <a href="http://translator.microsoft.com">http://translator.microsoft.com</a> .
Microsoft Translator Hub	<p>Центр Microsoft Translator дозволяє підприємствам та постачальникам мовних послуг створювати власні системи перекладу, які розуміють термінологію для бізнесу та галузі. Центр також може використовуватися разом із STF, дозволяючи адміністраторам затверджувати результати STF та додати їх безпосередньо до Hub. Microsoft Translator Hub доступний лише для статистичного машинного перекладу і не може використовуватися з новітньою версією API Microsoft Translator.</p> <p>Центр також використовується для збереження мови, дозволяючи громадам створювати власні мовні системи перекладу для збереження</p>

Назва функції	Опис
	мови та культури [24]. Центр був використаний для створення систем перекладу для таких мов, як Хмонг, Майя, Непалі та Валлійська.
Багатомовний інструментарій додатків (MAT)	Інструмент багатомовного додатку (MAT) - це інтегрований інструмент Visual Studio, який дозволяє розробникам оптимізувати робочі процеси локалізації своїх Windows, Windows Phone та настільних додатків. MAT покращує локалізацію керування файлами, підтримку перекладу та інструменти редагування.

#### 1.4 Висновки

Проаналізувавши існуючі рішення бачимо декілька тенденцій. По-перше кожна система перекладу створена з певною ціллю, ідеальна для виконання конкретних потреб та не є універсальною.

По-друге бачимо що кожного року з'являються нові розробки методів, алгоритмів для перекладу, покращуються алгоритми для збору паралельних корпусів текстів. Тема не є вивченою до кінця, ідеальний перекладач ще не створено, є багато мовних пар, що не підтримуються або підтримуються погано (якість перекладу незадовільна або переклад робиться через іншу мову з мовної групи транзитивно).

В цілому існує багато рішень для мовного перекладу з багатьма методами перекладу, але водночас можна знайти нішу для свого продукту, якщо він доведе свою конкурентоспроможність.

## 2 ПОНЯТТЯ МАШИННОГО ПЕРЕКЛАДУ

### 2.1 Вступ

Розглянемо систему машинного перекладу, засновану на комп'ютерному підході. На рисунку 1.1 зображено класичний трикутник Вокуа, який наочно демонструє рівні машинного перекладу таких систем та сам процес [25].



Рисунок 2.1 – Трикутник Вокуа [25].

Кожен з рівнів трикутника при реалізації є сам по собі складним компонентом. Внизу, на найнижчому, першому рівні лежить найпростіший метод машинного перекладу – послівне заміщення. Через те, що одне й те саме слово може бути перекладено у багатьох варіаціях цей підхід є дуже недосконалим – власне, проблема лексичної багатозначності на даному рівні є невирішеною. При цьому перестановка слів у реченні дуже часто являється причиною помилок при перекладі [26].

Глибший аналіз забезпечується на синтаксичному рівні, де береться до уваги зв'язок між словами. Хоча вирішити проблему багатозначності слів це допомагає

лише в тому випадку, якщо з урахуванням контексту можливе вживання лише одного варіанту – наприклад, коли кожне зі значень є іншою частиною мови.

Семантичний аналіз та трансфер виконується на рівні інтерлінгва – всеохоплююче представлення значення. Тож, чим вище ми піднімаємось до вершини трикутника Вокуа, тим ширше стає потенціал якості системи перекладу. Розроблювані системи по-різному інтерпретують цей трикутник.

На рисунку 2.2 проілюстровано процес будь-якого типу автоматичного перекладу. У цьому трикутнику зображений процес перетворення вихідного речення в цільове трьома різними шляхами.

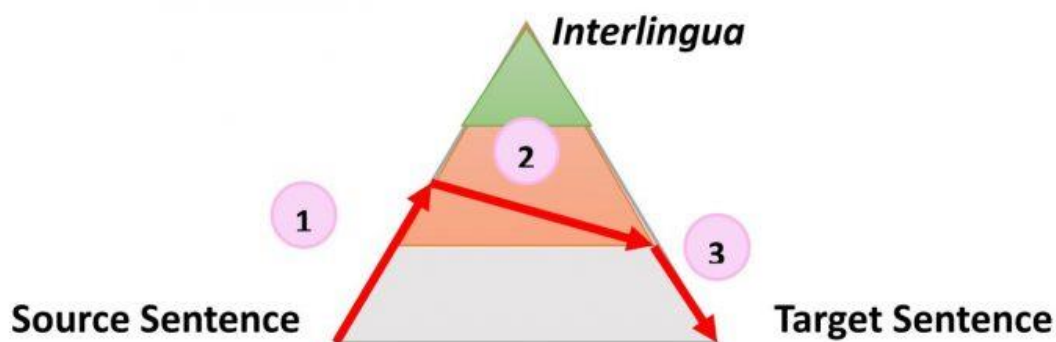


Рисунок 2.2 – Типовий процес автоматичного перекладу за Вокуа [27].

Ліва частина трикутника характеризує вихідну мову, коли як права – цільову. Різниця в рівнях всередині трикутника представляє глибину процесу аналізу вихідного речення, наприклад синтаксичного або семантичного. Тепер ми знаємо, що не можемо окремо проводити синтаксичний або семантичний аналіз, але теорія полягає в тому, що ми можемо заглибитися у кожному з цих напрямків. Перша червона стрілка позначає аналіз речення мовою оригіналу. З даного нам речення, яке є просто послідовністю слів, ми зможемо отримати уявлення про внутрішню структуру і ступені можливої глибини аналізу.

Наприклад, на одному рівні ми можемо визначити частини мови кожного слова (іменник, дієслово і тому подібне), а на іншому – взаємодія між ними. Наприклад, яке саме слово чи фраза є присудком.

Коли аналіз завершено, речення «переноситься» другим процесом з рівною або меншою глибиною аналізу на мову перекладу. Потім третій процес, який має назву «генерація», формує фактичне цільове речення з цієї інтерпретації, тобто створює послідовність слів цільовою мовою. Ідея використання трикутника полягає в тому, що чим вище (глибше) ви аналізуєте вихідне речення, тим простіше проходить фаза переносу. В кінцевому підсумку, якби ми могли перетворити вихідну мову в якийсь універсальний «інтерлінгвізм» під час цього аналізу, нам взагалі не потрібно було б виконувати процедуру перенесення. Знадобився б тільки аналізатор і генератор для кожної перекладеної мови на будь-яку іншу мову [28].

Ця загальна ідея і пояснює проміжні етапи, коли машина переводить речення покроково. Що ще більш важливо, ця модель описує характер дій під час перекладу.

Розглянемо методи роботи декількох технологій перекладу: технологію перекладу на базі правил, технологією фреймового перекладу та технологією перекладу на базі нейронних мереж, а також методи, які використовуються в кожному окремому випадку. Далі ми познайомимося з деякими прикладами і порівняємо, що кожна з технологій робить для того, щоб видати максимально правильний переклад.

## 2.2 Машинний переклад на базі правил

Машинний переклад на базі правил є найстарішим підходом (після перекладу на базі прямого перекладу, що сам по собі не являється методом) та охоплює найрізноманітніші технології. Однак, в основі всіх них зазвичай лежать наступні постулати [29]:

- процес строго слідує трикутнику Вокуа, аналіз дуже часто завищений, а процес генерації зводиться до мінімального;
- всі три етапи перекладу використовують базу даних правил та лексичних елементів, на які поширюються ці правила;

- правила та лексичні елементи задані однозначно, але можуть бути змінені лінгвістом.

Рівні аналізу:

- кожному слову привласнюється своя «частина мови», яка є граматичною категорією;
- морфологічний аналіз: слово «plays» розпізнається як спотворення від третьої особи і представляє форму дієслова «play»;
- семантичний аналіз: деяким словам присвоюється семантична категорія, наприклад, «violin» - інструмент;
- складовий аналіз: деякі слова згруповані. «Smart mouse» - це іменник;
- аналіз залежностей: слова і фрази пов'язані з «посиланнями», за допомогою яких відбувається ідентифікація об'єкта і суб'єкта дії основного дієслова «plays».

Перенесення такої структури речення буде підпорядковане таким правилам лексичного перетворення. Наприклад правила генерації французькою матимуть наступний вигляд:

- прикметник, виражений словосполученням, слідує за іменником — з декількома перерахованими винятками;
- визначальне слово узгоджено за кількістю та родом з іменником, який воно модифікує;
- прикметник погоджено по числу і статі з іменником, який воно модифікує;
- дієслово узгоджено з підметом.

Одним з методів перекладу у рамках перекладу на базі правил є переклад на базі трансферу. Трансферні системи застосовують до вхідного тексту правила, які ставлять у відповідність структури вихідної і цільової мов. Початковий етап роботи такої системи включає в себе аналіз тексту для створення формальної внутрішньої репрезентації тексту [30]. Переклад робиться на основі цієї репрезентації з використанням словників між двома мовами та використанням правил граматики. Внутрішнє уявлення та представлення тексту може мати декілька ступенів абстракції

для налаштування акцентів на важливих фрагментах тексту. Наприклад, в російській мові існує 6 загальновживаних відмінків, але взагалі – понад п'ятдесят [15 пз жук].

Один із можливих варіантів підходу на базі правил є інтерлінгвістичний підхід [31]. Системи, побудовані за цим підходом вважають третім поколінням систем машинного перекладу. Такі системи вивчає інтерлінгвістика як окрема наукова галузь. Вона має на меті досягнення лінгвістичної однорідності у всьому світі. Оригінальний текст трансформується в абстракцію оформлену у певний вигляд, що, на відміну від трансферного перекладу, не залежить від конкретної використовуваної мови. Перекладний текст створюється на основі цієї репрезентації.

Проте, на даний час реалізацій такого типу систем, які працювали б правильно хоча б для двох мов, не існує. Багато експертів висловлюють сумніви в можливості їх реалізації, принаймні на даному етапі розвитку людства [32]. При цьому культурні відмінності країн дуже сильно заважають створенню такої системи. Оскільки для створення такого роду систем необхідним є проектування міжмовного подання, яке повинно бути одночасно абстрактним і незалежним від кожної конкретної мови, але при цьому адекватно відображати її особливості. У рамках завдання побудови штучного інтелекту, все ще є завдання зробити так, щоб нейромережа «розуміла» текст. З урахуванням останніх розробок це стає можливим, але задача залишається не вирішеною [33].

## 2.3 Статистичний машинний переклад

Машинний переклад на базі фраз — це найпростіша й найпопулярніша версія статистичного машинного перекладу. Сьогодні він як і раніше є основним інструментом і використовується в великих онлайн-сервісах з перекладу.

Спершу розглянемо переклад заснований на прикладах — один з підходів до машинного перекладу, при якому використовується двомовний корпус тексту. В цьому полягає схожість з трансферними системами, де необхідна побудова мовних пар, проте використовувані механізми є зовсім різними.

Двомовний корпус під час перекладу використовується як база знань. Передбачається, що люди розкладають вихідний текст на фрази, потім перекладають ці фрази, а далі складають перекладний текст з фраз. Тобто мовні пари – тексти, що містять речення обома мовами, з/на які виконуватиметься переклад, можуть бути як варіантами написання двох пропозицій людиною – носієм двох мов, так і набором речень та їх перекладів, виконаних професійним перекладачем [34]. Для уникнення неоднозначності переклад фраз зазвичай відбувається за аналогією з попередніми перекладами. Наприклад, переклад російського слова «вертолёт» на українську мову як «вертоліт», «гвинтокрил» та «гелікоптер» для громадян України, які володіють двома мовами, цілком очевидний. Водночас, навіть суто переклад слова в такому випадку вводить елемент неоднозначності для СМП. За умови використання різноманітних синтаксичних та граматичних структур проблема набуває ще більшого масштабу. Водночас, усунення «зайвих» еквівалентних мовних структур беззаперечно приведе до збіднення мови.

Машинний переклад на базі фраз не слідує процесу, сформульованому Вокуа. Окрім того, в процесі цього типу машинного перекладу не проводиться ніякого аналізу або генерації, але, що більш важливо, додаткових частина не є детермінованою. Це означає, що технологія може генерувати кілька різних перекладів одного і того ж речення з одного й того ж джерела, а суть підходу полягає у виборі найкращого варіанту. Приблизний процес перекладу такої системи за трикутником Вокуа показано на рисунку 2.3.

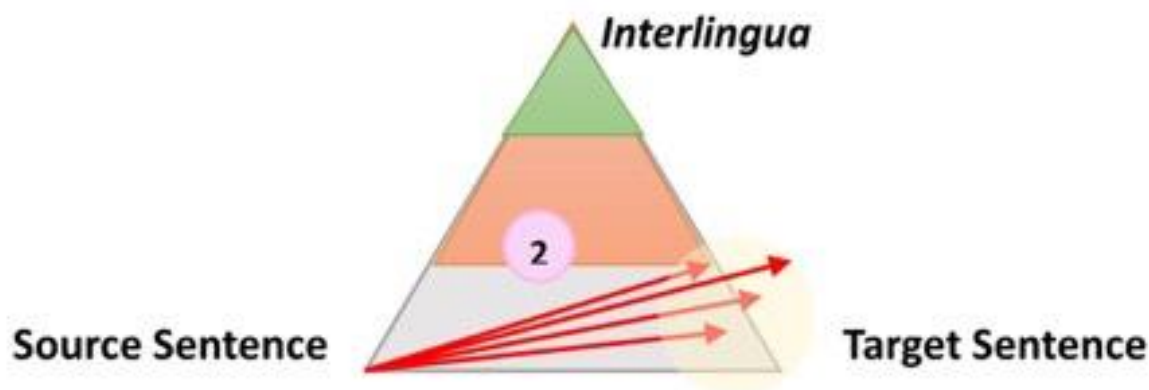


Рисунок 2.3 - Приблизний процес перекладу PBMT [26].



Модель перекладу РВМТ заснована на трьох базових методах:

- використання фрази-таблиці, яка дає варіанти перекладу і ймовірність їх вживання в цій послідовності мовою оригіналу;
- таблиця зміни порядку, яка вказує, як можуть бути переставлені слова при переносі з вихідної на мову перекладу;
- мовна модель, яка показує ймовірність для кожної можливої послідовності слів на цільову мову.

Двомовний корпус у якому речення до речення співставленні переклади двох мов під час перекладу використовується як база знань. Використовується припущення, що люди розбивають текст на речення, речення на фрази, і після цього люди перекладають ці фрази. Потім складають цільовий текст з фраз. Тобто мовні пари це тексти, які мають речення обома мовами, з яких або на які виконуватиметься переклад. Ці мовні пари можуть бути як написані звичайною людиною, так і набором речень та їх перекладів, виконаних професійним перекладачам. Для уникнення неоднозначності переклад фраз зазвичай відбувається за аналогією з попередніми перекладами. Наприклад, переклад російського слова «вертолёт» на українську мову як «вертолїт», «гвинтокрил» та «гелїкоптер» цілком очевидний для громадян України. Водночас, навіть у такому випадку це ставить СМП у позицію невизначеності. За умови використання різноманітних синтаксичних та граматичних структур проблема набуває ще більшого характеру. Водночас, усунення непотрібних, зайвих або еквівалентних мовних структур беззаперечно приведе до збіднення мови та знищення культури.

Отже, за даним методом на базі вихідного речення буде побудована таблиця можливостей перекладу кожного слова. Далі з цієї таблиці генеруються тисячі можливих варіантів перекладу речення. Однак завдяки інтелектуальним обчисленням ймовірності і використання досконалих алгоритмів пошуку, буде розглянуто тільки найбільш ймовірні варіанти перекладу, а кращий буде використано в якості підсумкового [35].

Пошукові алгоритми інтуїтивно вважають за ціллю використовувати послідовності слів, які є найбільш ймовірними перекладами вихідних з урахуванням

таблиці зміни порядку слів у реченні. Це дозволяє з високою точністю генерувати правильну послідовність слів на цільову мову.

У цьому підході немає явного або неявного лінгвістичного або семантичного аналізу. Нам було запропоновано безліч варіантів. Деякі з них краще, інші - гірше, але, наскільки відомо, основні онлайн-сервіси перекладу використовують саме цю технологію, навіть після появи перекладу на базі нейронних мереж.

## 2.4 Машинний переклад на базі нейронних мереж

Підхід до організації нейронного машинного перекладу кардинально відрізняється від попереднього і, спираючись на трикутник Вокуа, його можна описати графічно як показано на рисунку 2.4.

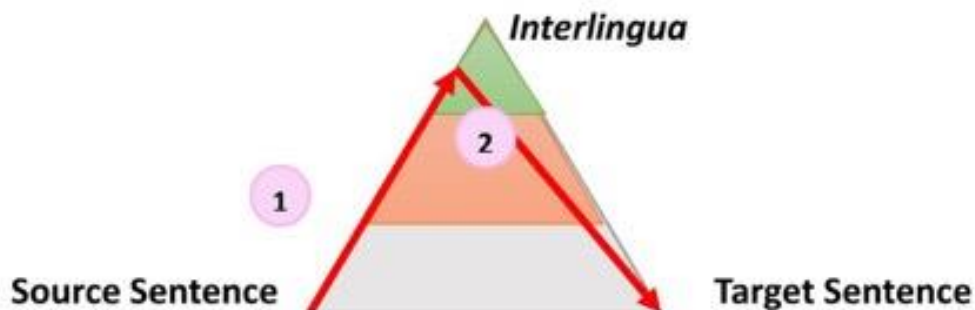


Рисунок 2.4 – Приблизний процес перекладу на базі нейронних мереж [26].

Нейронний машинний переклад має наступні особливості:

- «аналіз» називається кодуванням, а його результатом є загадкова послідовність векторів;
- «перенесення» називається декодуванням і безпосередньо генерує цільову форму без будь-якої фази генерації. Це не суворе обмеження і, можливо, є варіації, але базова технологія працює саме так [36].

Сам процес розбитий на дві фази. У першій кожне слово вихідного речення проходить через «кодер», який генерує те, що ми називаємо «вихідним контекстом», спираючись при цьому на поточне слово і попередній контекст. Структура етапу кодування показано на рисунку 2.5.

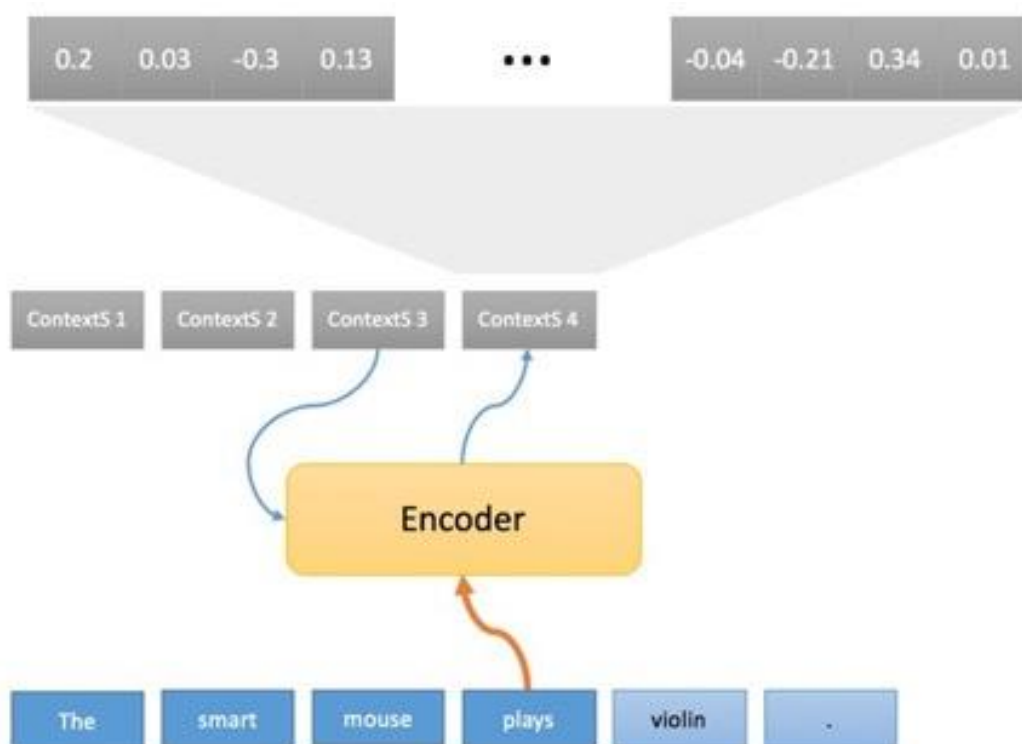


Рисунок 2.5 – Структура етапу кодування [37]

Послідовність вихідних контекстів (ContextS 1, ... ContextS 5) являє собою внутрішню інтерпретацію вихідного речення по трикутнику Вокуа і, як згадувалося вище, представляє з себе послідовність чисел з плаваючою комою (зазвичай тисяча чисел з плаваючою комою, пов'язаних з кожним вихідним словом). Поки ми не будемо обговорювати, як кодер виконує це перетворення, але хотілося б відзначити, що особливо цікавим є початкове перетворення слів у векторі «float».

Насправді це технічний блок, як і в випадку з заснованої на правилах системою перекладу, де кожне слово спочатку порівнюється зі словником, першим кроком кодера є пошук кожного вихідного слова всередині таблиці [37].

Припустимо, що вам потрібно уявити різні об'єкти з варіаціями за формою і кольором в двовимірному просторі. При цьому об'єкти, що знаходяться найближче один до одного повинні бути схожі. Уявімо собі двовимірну декартову площину, на якій розміщено фігури. На осі абсцис ми будемо намагатися помістити такі фігури, що мають найтісніший контакт за цим параметром об'єкти іншої форми (нам потрібно буде вказати, що робить фігури схожими. По осі ординат будемо розташовувати колір — зелений між жовтим і синім (розташовано так, тому що зелений є результатом змішання жовтого і синього кольорів). Якби у наших фігур були різні розміри, ми б могли додати цей третій параметр у вигляді додавання ще одного виміру. Якщо ми додамо більше кольорів або фігур, ми також зможемо збільшити і кількість вимірювань, щоб будь-яка точка могла представляти різні об'єкти і відстань між ними, яке відображає ступінь їх схожості.

Основна ідея в тому, що це працює і в разі розміщення слів. Замість фігур є слова, простір набагато більше – наприклад, ми використовуємо 800 вимірювань, але ідея полягає в тому, що слова можуть бути представлені в цих просторах з тими ж властивостями, що і фігури.

Отже, слова, що володіють загальними властивостями і ознаками будуть розташовані близько одне до одного. Наприклад, можна уявити, що слова певної частини мови — це один вимір, слова за ознакою статі (якщо такий є) — інший, може бути ознака позитивності або заперечності значення і так далі.

Другий крок перекладу має вигляд показаний на рисунку 2.6. На цьому етапі формується повна послідовність з нахилом «вихідний контекст», після чого один за іншим цільові слова генеруються з використанням:

- «цільового контексту», сформованого в зв'язці з попереднім словом і надає деяку інформацію про стан процесу перекладу;
- значущості «контекстного джерела», який являє собою суміш різних «вихідних контекстів» спираючись на конкретну модель під назвою «Модель уваги» (Attention Model). «Моделі уваги» вибирають вихідне слово для використання в перекладі на будь-якому етапі процесу;

- раніше наведеного слова з використанням вкладення слів для перетворення його в вектор, який буде оброблятися декодером.

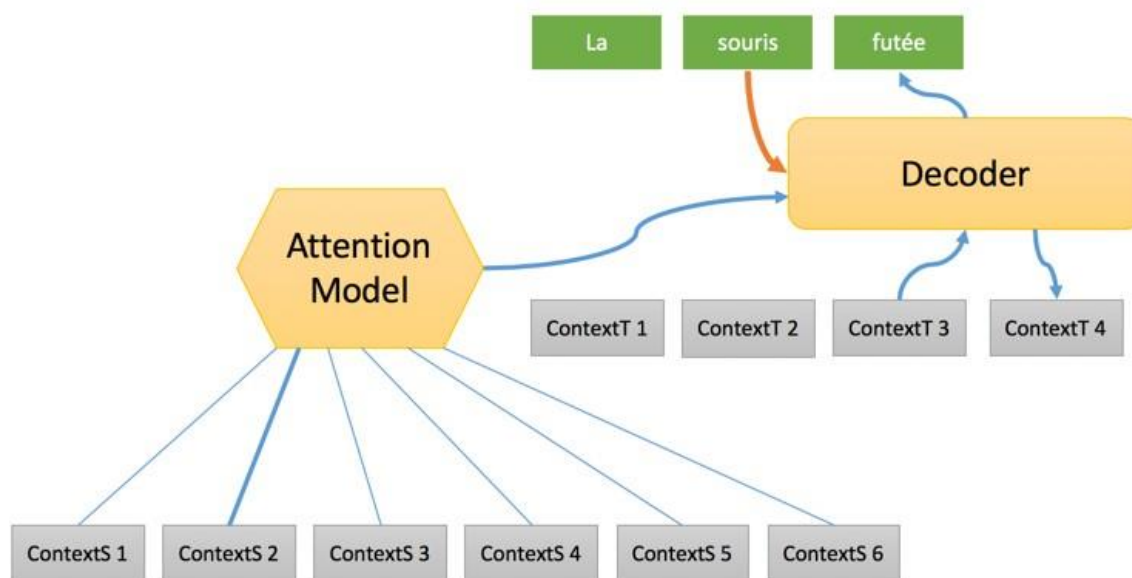


Рисунок 2.6 – Структура етапу декодування [37]

Переклад завершується, коли декодер доходить до етапу генерації фактично останнього слова в реченні. Можемо зробити висновок, що операції процесу нейронного машинного перекладу збудовані в тій же послідовності, що і в разі машинного перекладу на базі правил, однак характер операцій і обробка об'єктів повністю відрізняється. І починаються ці відмінності з перетворення слів в вектори через їх вкладення в таблиці.

## 2.5 Гібридні системи перекладу

Слабкою стороною систем, заснованих на правилах (RBMT) є лексичний вибір у трансфері, і відсутність стійкості у випадках помилок під час аналізу речень. Але в цілому їх переклад більш точний, без втрат вхідної інформації. Статистичні системи найбільш стійкі і завжди дозволяють отримати переклад, проте без гарантій якості перекладу. Ще вони мають складності у перекладі мовних феноменів, які потребують спеціальних лінгвістичних знань, таких як морфологія, синтаксис, порядок слів.

Саме ці недоліки були причиною появи мотивації для створення і розвитку гібридних систем. Гібридні системи – це системи які поєднують у собі підходи різних типів систем, вбираючи в себе їх переваги і компенсуючи недоліки [38].

Зараз виділяють три архітектурні наступні гібридних систем:

- поєднуючі системи (послідовні і паралельні): у таких системах для перекладу використовується дві або більше різних систем перекладу. Поєднання може бути або послідовним (одна система покращує переклад іншої) або паралельним (найкращий переклад вибирається з виходу двох систем). Системи з моделлю паралельного поєднання використовують декілька систем одночасно і мають механізм для вибору чи створення кращого перекладу на виході двох систем;
- архітектурні додатки – інтегрування нових компонентів у RBMT або SMT системи. У таких системах RBMT або SMT системи модифікуються за допомогою включення ресурсів відповідно іншої системи. Модифікації можуть бути двох типів: початкове редагування, модифікування ядра системи;
- справжні гібридні системи, що поєднують у собі компоненти різних систем. Вони використовують три основні компоненти:
  - ідентифікація частин вхідної мови;
  - переклад ідентифікованих частин на цільову мову використовуючи двомовні ресурси;
  - генерація речення цільової мови.

Алгоритм роботи типової гібридної системи:

- створення термінологічного словника з паралельних текстів для RBMT автоматичним шляхом;
- породження всіх можливих варіантів перекладу на основі:
  - лексичних варіантів;
  - варіантів синтезу різних конструкцій;
  - застосування постредагування.
- вибір найкращого варіанта, через реалізовану мовну модель [39].

## 2.6 Висновки

Всі подібні системи мають подібні переваги і недоліки. Розглянемо їх. Системи на базі правил поширені дуже сильно. Найбільш відомими представниками є Systran, Promt, ImTranslator. Зараз наявні платформи Apertium GramTrans для перекладу на базі правил [40].

Переваги:

- висока якість перекладу (забезпечується не методом як таким, а наявністю вивічених правил та баз даних);
- вибір галузевої направленості тексту, який підвищує якість перекладу;
- можливе уточнення перекладу, завдяки внесенню змін до словників перекладача (таким чином, користувач отримує потенційно нескінченну множину термінів, з якими можна вільно оперувати).

Недоліки:

- висока вартість та довготривала розробка;
- додавання нової мови вимагає повторного написання нової підсистеми для мови;
- потрібна команда кваліфікованих лінгвістів, для опису кожної з мов, а також для роботи уже з існуючими парами мов;
- необхідність у постійному оновленні лінгвістичних баз даних;
- неможливість змінити архітектуру систему у випадку, якщо вона виявиться невдалою;
- вимогливість до обчислювальних ресурсів на етапі створення бази, та неможливість використання поки система не навчилася перекладати до кінця.

Інтерлінгвістичні системи перекладу так і не були доведені до рівня промислових систем, їх характеристики зазначені на основі передбачень так припущень. Переваги:

- висока якість перекладу для будь-якої мови;

- універсальність для будь якої мови;
- масштабованість.

#### Недоліки:

- невідомо чи можна в повній мірі реалізувати дану систему;
- дуже важко й довго розробляти;
- висока ймовірність отримати неоптимізовану модель.

Системи машинного перекладу, засновані на статистичному перекладі, так само не мають яскравих представників. Існуючі відкриті прототипи мають чисто академічне значення для ілюстрації самого методу. Часто вони виглядають не як готовий продукт, який можна використовувати, а як набір бібліотек або методів, якими студенти або інші охочі можуть оперувати для навчальних цілей, дослідницьких цілей або створення своїх повноцінних продуктів під тою чи іншою ліцензією. Статистичні системи перекладу історично активно розроблялися компанією IBM (моделі перекладу IBM Model 1-5). Але найбільшу популярність цей метод набув завдяки Google SMT. Дуже часто промислові варіанти реалізації засекречені з чого можна зробити висновок про ефективність цього підходу для перекладу [41].

#### Переваги:

- висока якість перекладу (пропорційна натренованості);
- добре справляється з багатьма завданнями, залежними від контексту за допомогою статистичних методів;
- простота моделі;
- масштабованість;
- можливе навчання системи під час її експлуатації.
- побудова такої моделі це задача програміста, а не лінгвіста;
- таку систему можливо навчати під час використання.

#### Недоліки:

- для навчання необхідні якісні паралельні корпуси текстів які розміщені речення до речення;



- складний математичний апарат;
- якість перекладу прямо залежить від того, наскільки якісні корпуси тексту, та якість перекладу залежить від стилю перекладу;
- тривалий час навчання;
- нестабільність і непередбачуваність;
- велика кількість граматичних помилок;
- вимогливість до ресурсів на етапі навчання.

Системи на базі нейронних мереж останнім часом набули широкого використання за рахунок можливості наблизити модель перекладу до моделі яку використовує людина в своєму мозку для перекладу. Але це накладає свої певні обмеження та має свої переваги [42].

#### Переваги:

- висока якість перекладу у порівнянні з іншими системами;
- більше можливостей для того щоб долати проблеми багатозначності;
- відсутність машинного акценту при глибокому навчанні;
- високе “розуміння” контексту;
- можливість показувати хорошу якість перекладу навіть у тому випадку коли паралельний словник має речення, які перекладені не послівно а з використанням власних вражень, жаргонів, локальних або культурних фраз, тощо.

#### Недоліки:

- система на базі нейронних мереж потребує дуже багато часу для навчання;
- для навчання потрібен великий обсяг паралельних корпусів текстів;
- швидкість та якість перекладу залежить від обраної архітектури нейромережі.

Використання систем машинного перекладу що містять у собі дві або більше систем іншого типу, тобто гібридних систем має у собі багато переваг та долає багато недоліків, але такі системи самі по собі незрозумілі у виконанні та складні для цілей масштабування.

### Переваги:

- поєднання переваг різних систем перекладу;
- можливість подолати недоліки систем перекладу при правильному налаштуванні;
- висока стабільність перекладу;
- можливість прибрати машинний акцент перекладу.

### До недоліків таких систем можна віднести:

- відносна важкість виконання та підтримки системи;
- незрозумілий шлях масштабування;
- потреба у поєднанні декількох систем перекладу одночасно;
- можлива втрата продуктивності системи у певній мірі;
- неможливість легко сумістити декілька модулів, написаних на різних мовах програмування;
- потреба у рекомендаціях та супроводі лінгвістів для створення даної системи.

### 3 ФОРМУВАННЯ ВИМОГ ДО СИСТЕМИ ТА ПРОЕКТУВАННЯ

#### 3.1 Формування вимог до системи

Функціональні вимоги описані у таблиці 3.1.

Таблиця 3.1 - Функціональні вимоги

Шифр вимоги	Вимога	Опис вимоги
FR1	Німецько-український переклад тексту	Користувач може обрати напрямок перекладу з української на німецьку або навпаки. Користувач має змогу ввести текст в поле вводу на необхідній мові, а в сусідньому полі вводу отримати переклад після натискання відповідної клавіші.
FR2	Переклад веб-сторінки	Користувач має можливість виконати переклад сайту. Для цього треба ввести текст посилання в поле, де користувач також має змогу ввести текст для перекладу. Система має розпізнати гіперпосилання та автоматично згенерувати переклад сторінки у 2 форматах. Перший варіант являє собою повну версію перекладу, що означає, що кожен текст та слово, що є на сторінці перекладається і на вихід система має видати файл з розширенням .html. Файл має бути представлений у такому вигляді, щоб користувач міг використовувати сайт так, якби він був на оригінальному сайті з

Шифр вимоги	Вимога	Опис вимоги
		<p>можливістю переходів та ін. але з перекладом.</p> <p>Другий варіант являє собою варіант генерації файлу на виході, оптимізованого для читання, тобто на виході система має видавати .txt файл без виняткового форматування.</p>
FR3	Переклад файлів з розширенням .txt або .docx	<p>Користувач має змогу обрати файл для перекладу з розширенням .txt або .docx.</p> <p>Користувач має отримати на виході файл з аналогічним розширенням зі збереженням форматування та стилів. У випадку звичайного текстового файлу в ньому перекладаються речення, а всі відступи зберігаються. У випадку вибору перекладу .docx файлу система видає на виході такий самий файл з приміткою “_translated” перед розширенням файлу зі збереженням форматування, відступів та стилів. У файлі мають бути перекладені всі параграфи та таблиці.</p>
FR4	Історія перекладів	<p>Користувач має можливість переглянути історію своїх перекладів по трьом категоріям: текст, веб-сторінки, файли. Користувач має змогу обрати яка категорія буде показана в даний момент а яка ні.</p>

Шифр вимоги	Вимога	Опис вимоги
		<p>Історія має бути сформована в зворотному хронологічному порядку, що означає те, що останні елементи історії будуть нагорі. Історія має бути в основному вікні поруч з елементами вводу та виводу.</p> <p>Після успішного перекладу створюється новий елемент історії з яким користувач має змогу працювати. Також користувач має змогу видаляти елемент історії.</p> <p>У випадку текстового перекладу користувач має змогу виконати переклад повторно. У цьому випадку у поля вводу та виводу вставляються відповідні дані. Елемент історії має перейти у кінець історії якщо він був не в кінці.</p> <p>У випадку перекладу веб-сторінки створюється новий елемент історії з кнопками навігації. Користувач має мати змогу перейти на вихідний сайт за допомогою натискання, переходити до файлів повної або скороченої версії перекладу. Для цього мають бути відповідні кнопки у макеті елементу історії. Назви файлів на виході має містити у собі текст посилання веб-сторінки та примітку “_full_translation” та “_readable_translation”</p>

Шифр вимоги	Вимога	Опис вимоги
		<p>для повної версії та версії для читання відповідно. Має бути ще 2 варіанти взаємодії: користувач має змогу відразу відкрити без переходу повну та скорочену версію перекладу.</p> <p>У випадку перекладу файлу також створюється новий елемент історії з кнопками навігації. Користувач має змогу перейти до вихідного файлу за допомогою натискання або мати змогу одразу відкрити цей файл не переходячи до нього. Для цього також мають бути відповідні кнопки у макеті елемента історії. Файл має маркуватись “_translated”.</p>

Нефункціональні вимоги описані у таблиці 3.2.

Таблиця 3.2 – Нефункціональні вимоги

Шифр вимоги	Вимога
NFR1	Сумісність програми з операційними системами OS X 10.4.2+ та Windows 7+.
NFR2	Користувач має мати можливість перекласти текст до 1000 символів.
NFR3	Підсистема перекладу має перекладати текст розміром 1000 символів не більше 5 секунд.

Шифр вимоги	Вимога
NFR4	Користувач має бути оповіщений про всі дії які виконуються у програмі та наочно бачити статус виконання процесу.
NFR5	Усі іконки кнопок без текстової примітки мають бути доступні для наведення з ціллю отримання підказку про те яку дію або функцію забезпечує дана кнопка.
NFR6	Користувач має мати змогу розширювати та звужувати вікно програми, при цьому дизайн має автоматично підлаштовуватись до розміру вікна, яке задав користувач.
NFR7	Дизайн інтерфейсу користувача має бути виконаний в тонах, які не викликають подразнення, без різких для очей елементів та у кольоровій гаммі синього кольору.
NFR8	Наявність інформації про версію програмного забезпечення та номер її зборки.

### 3.2 Сценарії використання системи

Діаграма прецедентів або діаграма варіантів використання представлена у додатку Б. Матриця відстеження між вимогами та варіантами використання приведена у додатку В.

Кожен варіант використання системи представлений у таблиці 3.3.

Таблиця 3.3 – Варіанти використання системи

Шифр варіанту використання	Назва варіанту використання
UC-1	Німецько-український переклад
UC-1.1	Переклад файлу з розширеннями .txt або .docx
UC-1.2	Переклад тексту
UC-1.3	Переклад сайту
UC-1.3.1	Генерація файлів з результатами перекладу на виході
UC-2	Перегляд історії перекладу
UC-2.1	Взаємодія з елементами історії
UC-2.1.1	Перехід до згенерованих файлів перекладу
UC-2.1.2	Відкриття згенерованих файлів перекладу
UC-2.1.3	Повтор перекладу
UC-2.1.4	Перехід по гіперпосиланню на перекладену веб-сторінку
UC-2.2	Сортування
UC-2.2.1	Показувати історію перекладу файлів
UC-2.2.2	Показувати історію перекладу сайтів
UC-2.2.3	Показувати історію перекладів текстів
UC-2.3	Видалення елемента історії
UC-4.1	Спрощеного текстового файлу варіанту веб-сторінки для читання
UC-4.2	Повного html файлу

Приведемо опис кожного варіанту використання у таблицях 3.4-3.21. За замовчуванням будемо вважати що суб'єкти варіантів використання це користувач системи та сама програма.

Опис варіанту використання за шифром UC-1 приведено у таблиці 3.4.



Таблиця 3.4 – Опис варіанту використання UC-1

Шифр	UC-1
Ім'я	Німецько-український переклад
Опис	Користувач має змогу здійснити переклад контенту між українською та німецькою мовами
Передумови	Обрано контент для перекладу
Результат	Отримано перекладений контент та створено новий елемент історії
Основний потік	<ol style="list-style-type: none"> <li>1. Користувач налаштовує напрямок перекладу</li> <li>2. Користувач вказує контент для перекладу</li> <li>3. Натискає на кнопку перекладу</li> </ol>
Виключення	<ol style="list-style-type: none"> <li>1. Контент не введено</li> <li>2. Контент невірної формату</li> </ol>

Опис варіанту використання за шифром UC-1.1 приведено у таблиці 3.5.

Таблиця 3.5 – Опис варіанту використання UC-1.1

Шифр	UC-1.1
Ім'я	Переклад файлу з розширеннями .txt або .docx
Опис	Користувач обирає файл для перекладу з розширенням .txt або .docx. і отримує на виході файл з аналогічним розширенням зі збереженням форматування та стилів. У випадку звичайного текстового файлу в ньому перекладаються речення, а всі відступи зберігаються. У випадку вибору перекладу .docx файлу система видає на виході файл зі збереженням форматування, відступів та стилів. У файлі мають бути перекладені всі параграфи та таблиці.
Передумови	Обрано файл та натиснута кнопка «Перекласти»

Результат	<ol style="list-style-type: none"> <li>1. Файл з аналогічним розширенням та зі збереженням форматування та стилів.</li> <li>2. Створено елемент історії</li> </ol>
Основний потік	<ol style="list-style-type: none"> <li>1. Користувач натискає на кнопку обирання файлу для перекладу</li> <li>2. Користувач обирає файл з спливаючого вікна</li> </ol>
Виключення	<ol style="list-style-type: none"> <li>1. Файл невірної формату</li> </ol>

Опис варіанту використання за шифром UC-1.2 приведено у таблиці 3.6.

Таблиця 3.6 – Опис варіанту використання UC-1.2

Шифр	UC-1.2
Ім'я	Переклад тексту
Опис	Користувач отримує перекладений текст на виході
Передумови	Введено текст та натиснута кнопка перекладу
Результат	<ol style="list-style-type: none"> <li>1. У полі для результату перекладу з'являється перекладений текст</li> <li>2. Створено елемент історії</li> </ol>
Основний потік	<ol style="list-style-type: none"> <li>1. Користувач обирає напрямок перекладу</li> <li>2. Користувач вводить в поле для перекладу текст</li> <li>3. Натискає на кнопку «Перекласти»</li> </ol>
Виключення	<ol style="list-style-type: none"> <li>1. Текст для перекладу не введено</li> <li>2. Текст для перекладу перевищує 1000 символів</li> </ol>

Опис варіанту використання за шифром UC-1.3 приведено у таблиці 3.7.

Таблиця 3.7 – Опис варіанту використання UC-1.3

Шифр	UC-1.3
Ім'я	Переклад сайту

Опис	Користувач має змогу обрати сайт для перекладу
Передумови	Обрано сайт для перекладу. Присутній інтернет.
Результат	На виході маємо 2 файли для перекладу описані у варіанті використання UC-1.3.1
Основний потік	<ol style="list-style-type: none"> <li>1. Користувач вводить веб-посилання в текстове поле для перекладу</li> <li>2. Натискає на кнопку «Перекласти»</li> </ol>
Виключення	<ol style="list-style-type: none"> <li>1. Сайт не введено в текстове поле</li> <li>2. Помилка з'єднання</li> <li>3. Помилка доступу до сайту. Можливо сайт введено неправильно.</li> </ol>

Опис варіанту використання за шифром UC-1.3.1 приведено у таблиці 3.8.

Таблиця 3.8 – Опис варіанту використання UC-1.3.1

Шифр	UC-1.3.1
Ім'я	Генерація файлів з результатами перекладу на виході
Опис	При перекладі сайту користувач отримує спрощену версію перекладу для читання
Передумови	Описано у UC-1.3
Результат	На виході користувач має 2 файли з результатами перекладу веб-сторінки. Перший у спрощеному форматі для читання у текстовому форматі з розширенням .txt. Цей файл не є інтерактивним, а є лиш версією для ознайомлення з текстовим контентом сайту. Другий файл – це повний переклад веб-сторінки з розширенням html. Користувач має змогу відкрити файл у браузері та використовувати вихідну навігацію сайту але у перекладеній версії. Також створено новий елемент історії який розміщено на початку списку

Основний потік	Описано у UC-1.3
Виключення	Описано у UC-1.3

Опис варіанту використання за шифром UC-2 приведено у таблиці 3.9.

Таблиця 3.9 – Опис варіанту використання UC-2

Шифр	UC-2
Ім'я	Перегляд історії перекладу
Опис	Користувач має змогу переглянути історію перекладів
Передумови	Наявні переклади до моменту перегляду
Результат	Список історій перекладу
Основний потік	Користувач перекладає контент і має його змогу побачити у історії
Виключення	1. Помилка читання бази даних

Опис варіанту використання за шифром UC-2.1 приведено у таблиці 3.10.

Таблиця 3.10 – Опис варіанту використання UC-2.1

Шифр	UC-2.1
Ім'я	Взаємодія з елементами історії
Опис	Користувач має змогу натискати на кнопки взаємодії з елементами історії
Передумови	Наявні переклади до моменту перегляду
Результат	Результати показані в наступних варіантах використання
Основний потік	Користувач натискає кнопки на елементах історії
Виключення	Результати приведені в наступних варіантах використання

Опис варіанту використання за шифром UC-2.1.1 приведено у таблиці 3.11.

Таблиця 3.11 – Опис варіанту використання UC-2.1.1

Шифр	UC-2.1.1
Ім'я	Перехід до згенерованих файлів перекладу
Опис	Користувач має змогу перейти до файлів в файловій ієрархії
Передумови	<ol style="list-style-type: none"> <li>1. Наявний елемент історії з результатом перекладу у вигляді файлу.</li> <li>2. Користувач натиснув відповідну кнопку для переходу до файлу</li> </ol>
Результат	Відкривається менеджер файлів з виділеним файлом
Основний потік	Користувач натискає кнопку для переходу по файлу та бачить файл у файловому менеджері
Виключення	<ol style="list-style-type: none"> <li>1. Файл більше не існує. В цьому випадку відкривається менеджер файлів у папці, призначеної для зберігання результатів перекладу</li> </ol>

Опис варіанту використання за шифром UC-2.1.2 приведено у таблиці 3.12.

Таблиця 3.12 – Опис варіанту використання UC-2.1.2

Шифр	UC-2.1.2
Ім'я	Відкриття згенерованих файлів перекладу
Опис	Користувач має змогу одразу відкрити файл без переходу до файлу в файловій ієрархії
Передумови	<ol style="list-style-type: none"> <li>1. Наявний елемент історії з результатом перекладу у вигляді файлу.</li> <li>2. Користувач натиснув відповідну кнопку для відкриття відповідного файлу</li> </ol>

Результат	Файл відкрито з використанням системного способу, встановленого за замовчуванням та існуючого в одній з цільових ОС.
Основний потік	<ol style="list-style-type: none"> <li>1. Користувач ходить по історії перекладів</li> <li>2. При натисканні на відповідну кнопку програма відкриває обраний файл</li> </ol>
Виключення	<ol style="list-style-type: none"> <li>1. Файлу не існує</li> </ol>

Опис варіанту використання за шифром UC-2.1.3 приведено у таблиці 3.13.

Таблиця 3.13 – Опис варіанту використання UC-2.1.3

Шифр	UC-2.1.3
Ім'я	Повтор перекладу
Опис	Користувач має змогу перекласти текст ще один раз
Передумови	Наявність елементу історії
Результат	<ol style="list-style-type: none"> <li>1. Текст перекладено ще раз</li> <li>2. Створено новий елемент історії та поміщено його в початок списку.</li> <li>3. Елемент історії має стосуватись історії перекладу тексту</li> <li>4. Вихідний текст та перекладений текст розміщено у відповідних полях для вводу</li> </ol>
Основний потік	<ol style="list-style-type: none"> <li>3. Користувач ходить по історії перекладів</li> <li>4. При натисканні у будь-якому місці вільному від інших кнопок програма повторює переклад</li> </ol>
Виключення	Відсутні

Опис варіанту використання за шифром UC-2.1.4 приведено у таблиці 3.14.

Таблиця 3.14 – Опис варіанту використання UC-2.1.4

Шифр	UC-2.1.4
Ім'я	Перехід по гіперпосиланню на перекладену веб-сторінку
Опис	Користувач має змогу відкрити вихідну веб-сторінку у браузері
Передумови	<ol style="list-style-type: none"> <li>1. Наявний відповідний елемент історії</li> <li>2. Натиснуто на відповідну кнопку</li> </ol>
Результат	Веб-сторінку відкрито у браузері
Основний потік	<ol style="list-style-type: none"> <li>1. Користувач використовує навігацію по елементах історії</li> <li>2. Користувач натискає на кнопку переходу</li> </ol>
Виключення	Відсутні

Опис варіанту використання за шифром UC-2.2 приведено у таблиці 3.15.

Таблиця 3.15 – Опис варіанту використання UC-2.2

Шифр	UC-2.2
Ім'я	Сортування
Опис	У повсякденні користувач перекладає десятки елементів контенту щоденно, тож для того, щоб він мав можливість зручно використовувати програму він має мати змогу відфільтрувати елементи історії за типом перекладу: переклад тексту, переклад веб-сторінок та переклад файлів.
Передумови	<ol style="list-style-type: none"> <li>1. Наявні елементи історії</li> <li>2. Обрано відповідні фільтри</li> </ol>
Результат	Показуються тільки ті елементи історії, які дозволив показувати користувач за допомогою фільтрів у вигляді чек боксів.

Основний потік	<ol style="list-style-type: none"> <li>1. Користувач обирає фільтри для історії</li> <li>2. Деякі елементи історії не показуються</li> </ol>
Виключення	У випадку якщо елементи історії відсутні, то фільтри не мають ефекту, але при цьому програма не має входити в критичний стан

Опис варіанту використання за шифром UC-2.2.1 приведено у таблиці 3.16.

Таблиця 3.16 – Опис варіанту використання UC-2.2.1

Шифр	UC-2.2.1
Ім'я	Показувати історію перекладу файлів
Опис	Користувач бажає бачити в історії елементи історії які пов'язані з перекладом файлів
Передумови	Обрано відповідну кнопку
Результат	Історія містить в тому числі елементи історії які пов'язані з перекладом файлів
Основний потік	Користувач обирає відповідний важіль фільтрів
Виключення	Відсутні

Опис варіанту використання за шифром UC-2.2.2 приведено у таблиці 3.17.

Таблиця 3.17 – Опис варіанту використання UC-2.2.2

Шифр	UC-2.2.2
Ім'я	Показувати історію перекладу сайтів
Опис	Користувач бажає бачити в історії елементи історії які пов'язані з перекладом веб-сторінок
Передумови	Обрано відповідну кнопку
Результат	Історія містить в тому числі елементи історії які пов'язані з перекладом веб-сторінок



Основний потік	Користувач обирає відповідний важіль фільтрів
Виключення	Відсутні

Опис варіанту використання за шифром UC-2.2.3 приведено у таблиці 3.18.

Таблиця 3.18 – Опис варіанту використання UC-2.2.3

Шифр	UC-2.2.3
Ім'я	Показувати історію перекладів текстів
Опис	Користувач бажає бачити в історії елементи історії які пов'язані з перекладом файлів
Передумови	Обрано відповідну кнопку
Результат	Історія містить в тому числі елементи історії які пов'язані з перекладом текстів
Основний потік	Користувач обирає відповідний важіль фільтрів
Виключення	Відсутні

Опис варіанту використання за шифром UC-2.3 приведено у таблиці 3.19.

Таблиця 3.19 – Опис варіанту використання UC-2.3

Шифр	UC-2.3
Ім'я	Видалення елемента історії
Опис	Користувач може видалити елементи історії які йому більше не потрібні
Передумови	Наявний елемент в історії
Результат	Елемент історії видалено назавжди
Основний потік	1. Користувач обирає елемент історії для видалення 2. Користувач натискає кнопку для видалення елемента історії
Виключення	Відсутні

Опис варіанту використання за шифром UC-4.1 приведено у таблиці 3.20.

Таблиця 3.20 – Опис варіанту використання UC-4.1

Шифр	UC-4.1
Ім'я	Спрощеного текстового файлу варіанту веб-сторінки для читання
Опис	Цей варіант використання описує варіант файлу, який згенеровано на виході в результаті перекладу веб-сторінки. Даний файл являє собою скорочену версію перекладу веб-сторінки яка призначення для читання
Передумови	<ol style="list-style-type: none"> <li>1. Перекладено веб-сторінку</li> <li>2. Обрано перехід до файлу</li> <li>3. Обрано відкриття файлу</li> </ol>
Результат	Описано в попередніх варіантах використання
Основний потік	Описаний в попередніх варіантах використання
Виключення	Описані в попередніх варіантах використання

Опис варіанту використання за шифром UC-4.2 приведено у таблиці 3.21.

Таблиця 3.21 – Опис варіанту використання UC-4.2

Шифр	UC-4.2
Ім'я	Повного html файлу
Опис	Цей варіант використання описує варіант файл на виході який згенеровано в результаті перекладу веб-сторінки. Даний файл являє собою повну версію перекладу веб-сторінки з присутньою вихідною навігацією
Передумови	<ol style="list-style-type: none"> <li>1. Перекладено веб-сторінку</li> <li>2. Обрано перехід до файлу</li> <li>3. Обрано відкриття файлу</li> </ol>

Результат	Описано в попередніх варіантах використання
Основний потік	Описаний в попередніх варіантах використання
Виключення	Описані в попередніх варіантах використання

### 3.3 Розгортання системи та бізнес вимоги

Діаграма розгортання системи приведена у додатку Г. Розглянемо детальніше структуру розгортання. Для цього в таблиці 3.22 опишемо кожен елемент.

Таблиця 3.22 – Опис елементів діаграми розгортання

Шифр елемента	Назва елемента	Опис у контексті системи та вимоги
DD-1	Сервіси Microsoft	Сервіси Microsoft включають в себе підсистему для сховища додатків та підсистему для рекламних кампаній
DD-1.1	Microsoft Store	Це сервіс від Microsoft що дозволяє власнику додатку залити додаток на сервер Microsoft та дозволити мільйонам користувачів скачати додаток з перевіреного сервісу Microsoft Store, зробити перегляд відгуків, та інше. Використання такого підходу дозволяє не запускати свій хостинг сайту на якому буде зберігатися програма для різних версій систем. Також це збільшує у користувачів довіру до продукту, адже їм не треба турбуватись про безпеку сайту, за них це зробить Microsoft
DD-1.1.1	Сховище додатків	Це компонент системи Microsoft Store з яким користувач працює для того, щоб завантажити додаток на свій пристрій на базі операційної системи Windows

Шифр елемента	Назва елемента	Опис у контексті системи та вимоги
DD-1.2	Microsoft promotion service	Цей компонент являє собою можливість робити маркетингові кампанії для того щоб просувати додаток в межах Microsoft Store. Включає в себе такі методи просування як природня реклама при пошуку, статус перевіреного продавця, та ін.
DD-2	Сервіси Apple	Сервіси Apple включають в себе підсистему для сховища додатків та підсистему для рекламних кампаній
DD-2.1	App Store	Виконує схожі функції з Microsoft Store, але для користувачів операційної системи OS X
DD-2.1.1	Сховище додатків	Це компонент системи App Store з яким користувач працює для того, щоб завантажити додаток на свій пристрій на базі операційної системи OS X
DD-2.2	Search Ads Service	Цей компонент також дозволяє робити маркетингові кампанії, але в межах App Store. Дозволяє просувати додаток за допомогою реклами при пошуку
DD-3	Комп'ютер на базі Windows	Являє собою пристрій користувача, який користується системою Windows
DD-4	Комп'ютер на базі OS X	Являє собою пристрій користувача, який користується системою OS X
DD-5	Середовище Python	Середовище мови Python дозволяє не писати програму декілька разів для різних систем. В даному випадку підсистеми графічного інтерфейсу та перекладу пишуться тільки один раз,

Шифр елемента	Назва елемента	Опис у контексті системи та вимоги
		компілюються під різні платформи та зберігаються в Microsoft Store або в App Store для подальшого завантаження користувачем
DD-6	Додаток	Це система перекладу у вигляді файлу для запуску
DD-7	Адмін-панель	Адмін-панель панель являє собою компонент підсистем для того, щоб можна було користуватись функціями які надають описані підсистеми. Доступ до адмін-панелі має бути тільки у володаря додатку або у головного архітектора проекту

### 3.4 Компоненти системи

Діаграма компонентів наведена у додатку Г. Розглянемо її детальніше. Опис кожного компоненту наведено у таблиці 3.23.

Таблиця 3.23 – Опис компонентів системи

Назва компоненту	Опис
Ядро системи	Ядро системи відповідає за з'єднання всіх компонентів в одне ціле та підтримку їх роботи. Тримає у собі бізнес-логіку додатку
Модуль роботи зі сховищем	Модуль роботи зі сховищем призначений для створення, читання файлів з або на файлову систему пристрою за вказаною адресою та вказаними параметрами
Модуль інтерфейсу користувача	Модуль призначений для створення та вимальовування, зміну, роботу інтерфейсу користувача

Назва компоненту	Опис
	за допомогою обраної бібліотеки та обробку подій викликаних користувачем
Модуль роботи з інтернетом	Модуль роботи з інтернетом призначений для читання даних за вказаним посиланням, тобто зчитування html сторінок
Модуль перекладу	Модуль перекладу призначений для того, щоб перекладати текст за вказаними параметрами
Модуль роботи з лінгвістичною базою даних	Модуль роботи, який існує у складі модуля перекладу і призначений для того, щоб обробляти лінгвістичну базу даних за математичною моделлю перекладу, яка вказана у відповідному розділі
Модуль роботи з документами	Це модуль призначений для зчитування документів з розширеннями .docx та .txt, обробку даних в середині, зміну даних зі збереженням форматів та стилів та збереження файлів на виході у відповідному форматі

### 3.5 Обґрунтування вибору компонентів

Розглянуто різні компоненти та обґрунтовано вибір кожного у таблиці 3.24.

Таблиця 3.24 – Технологічна здійсненність ідеї проекту

п\п	Застосування технології	Технології реалізації
1		Kivy
2		Pyforms
3		PyGObject (a.k.a. PyGi)
4		PyQt

5		PyGUI
Обрана технологія реалізації користувацького інтерфейсу: PyQt. Ця бібліотека є найпростішою серед обраних бібліотек і є безкоштовною для використання. Всі можливості які надає ця бібліотека дозволяють виконати інтерфейс користувача у повному обсязі		
11		Miscrosoft SQL Server
12		ORACLE
13		PostreSQL
14		MongoDb
		SqLite
Обрана технологія реалізації ідеї проекту: SqLite. Ця СУБД є найпростішою для виконання взаємодії бази даних. Також ця СУБД займає менше місця на файловій системі та є частовживаною		
15		Python
16		Java
17		C#
18		ReactJs
15		C++
Обрана технологія реалізації ідеї проекту: Python. Середовище розробки Python дозволяє створити єдину базу коду для OS X та Windows операційних систем. Також на Python пишуться більшість бібліотек для перекладу, тому інтеграція модулю перекладу з ними буде максимально комфортною.		

### 3.6 Структура бази даних

Концептуальна модель бази даних SQLite приведена у додатку Д. Розглянемо її детальніше. У таблиці 3.25 описані поля таблиці Languages. Таблиця являє собою перелік мов та їх ідентифікатори.

Таблиця 3.25 – Поля таблиці Languages

id	Ідентифікатор
Name	Текстове представлення

У таблиці 3.26 описані поля таблиці TranslationTypes. Таблиця являє собою перелік типів елементів історії та статус дозволу їх показу користувачем на графічному інтерфейсі

Таблиця 3.26 – Поля таблиці TranslationTypes

id	Ідентифікатор
type_name	Текстове представлення
is_enabled	Статус відображення на інтерфейсі користувача

У таблиці 3.27 описані поля таблиці History. History описує базовий елемент історії.

Таблиця 3.27 – Поля таблиці History

id	Ідентифікатор
type_id	Індекс типу елементу історії. На момент написання є 3 типи історії: текстовий переклад, переклад веб-сторінки та переклад файлу. За допомогою цього поля маємо брати дані відповідно у таблицях TranslationTypeText, TranslationTypeWeb та TranslationTypeFile.
timestamp	Штам часу запису елементу історії в базу даних



translating_from_id	Індекс мови перекладу з якої ми перекладаємо. Посилається на індекс мови у таблиці Languages.
---------------------	-----------------------------------------------------------------------------------------------

У таблиці 3.28 описані поля таблиці TranslationTypeText. Таблиця відображає елемент історії перекладу тексту.

Таблиця 3.28 – Поля таблиці TranslationTypeText

id	Ідентифікатор який є ідентичним ідентифікатору з таблиці History
src_text	Запис тексту який перекладав користувач
trg_text	Запис результату перекладу

У таблиці 3.29 описані поля таблиці TranslationTypeFile. Таблиця відображає елемент історії перекладу файлу.

Таблиця 3.29 – Поля таблиці TranslationTypeFile

id	Ідентифікатор який є ідентичним ідентифікатору з таблиці History
src_filepath	Шлях до файлу який був обраний для перекладу
trg_filepath	Шлях до файлу який створився в результаті перекладу

У таблиці 3.30 описані поля таблиці TranslationTypeWeb. Таблиця відображає елемент історії перекладу веб-сторінки.

Таблиця 3.30 – Поля таблиці TranslationTypeWeb

id	Ідентифікатор який є ідентичним ідентифікатору з таблиці History
src_web_page	Текстове представлення веб-посилання на веб-сторінку яку обрав користувач для перекладу
trg_file_path_simple	Шлях до файлу результату перекладу веб-сторінки для читання
trg_file_path_full	Шлях до повної версії перекладу веб-сторінки

### 3.6 Розроблення інтерфейсу користувача

Макет інтерфейсу користувача приведено у додатку Е. Інтерфейс користувача містить вікно додатку з описаними елементами та контейнерами. Опис макету наведено у таблиці 3.31.

Таблиця 3.31 – Опис макету користувацького інтерфейсу.

Шифр елемента	Назва елемента	Опис
UI-1	Поле для вводу	Це поле для вводу тексту для перекладу або посилання на веб-сторінку яка має бути автоматично розпізнана програмою
UI-2	Поле для виводу	Поле у якому з'являється перекладений текст у випадку текстового перекладу.
UI-3	Кнопка зміни напрямку перекладу	Кнопка для того щоб контролювати напрямок перекладу з української на німецьку або з німецької на українську

Шифр елемента	Назва елемента	Опис
UI-4	Контейнер заголовку історії	Заголовок історії який містить у собі 3 статуси кнопки. Кожна з них може бути у 2 станах: натиснута або відтиснута. Натиснутий стан означає що користувач бажає бачити елемент історії. Кожна кнопка відповідає за свій тип елементу історії: текстовий переклад, переклад файлів та переклад веб-сторінок. При натисканні на кнопки контейнер елементів історії відображає актуальні фільтри користувача
UI-5	Контейнер елементів історії	Контейнер який містить елементи історії відповідних типів. Вони розташовується у зворотному хронологічному порядку – це означає, що останній переклад буде відображено угорі. У кожного елемента історії є кнопка видалення із списку після натискання на яку видаляється елемент з історії та з бази даних
UI-5.1	Контейнер елемента історії текстового перекладу	Тип елемента історії який показує переклад тексту. При натисканні будь-де на цьому контейнері робиться повторний переклад при якому у поля вводу та виводу вставляються відповідні дані та напрямок перекладу вказується такий, який був використаний при перекладі.
UI-5.2	Контейнер елемента історії перекладу веб-сторінки	Тип елемента історії який показує переклад веб-сторінки. При натисканні на гіперпосилання веб-сторінки програма має відкрити веб-сайт за допомогою системного інтерфейсу. Контейнер містить 4 кнопки. Дві з них дозволяють перейти у

Шифр елемента	Назва елемента	Опис
		файловому менеджері до згенерованих файлів перекладу, інші дві дозволяють одразу відкрити генеровані файли для перекладу. Існує 2 файли: версія для читання та повна версія перекладу.
UI-5.3	Контейнер елемента історії перекладу файлу	Тип елемента історії який показує переклад файлу. Контейнер містить 2 кнопки. Кнопка «Перейти» дозволяє перейти до файлу з відкриттям менеджера файлів за допомогою системного інтерфейсу. Кнопка «Відкрити» сигналізує програмі відкрити файл
UI-6	Кнопка перекладу	Кнопка яка створена для того щоб почати переклад обраного контенту
UI-7	Кнопка вибору файлу	Кнопка яка створена для того, щоб обрати файл за допомогою системного інтерфейсу для перекладу.
UI-7	Статус виконання	Статус виконання має вказувати на результат виконання операції незалежно від того він позитивний негативний. Наприклад «Файл перекладено», або «Файл невірною формату»

### 3.7 Тестування системи

Наведемо набір тестів (test suite) які покривають функціональні вимоги. Матриця відстеження між тест-кейсами та функціональними вимогами наведена у додатку Є. Усі тест кейси описано в таблиці 3.32.

Таблиця 3.32 – Опис тест кейсів

Шифр тест кейсу	Опис тест кейсу
ТС-1	<p>Назва: Переклад тексту</p> <p>Кроки:</p> <ol style="list-style-type: none"> <li>1. Набрати або вставити текст у поле вводу.</li> <li>2. Натиснути кнопку «Перекласти».</li> </ol> <p>Очікуваний результат: У полі виводу відображається переклад тексту. Також створено елемент історії перекладу.</p>
ТС-2	<p>Назва: Перевірка ліміту введення на текст</p> <p>Кроки: ввести в поле вводу більше 1000 символів</p> <p>Очікуваний результат: Додаток обрізає символи після 1000</p>
ТС-3	<p>Назва: Переклад файлу</p> <p>Кроки:</p> <ol style="list-style-type: none"> <li>1. Натиснути на кнопку вибору файлу для перекладу.</li> <li>2. Обрати файл зі спливаючого вікна</li> </ol> <p>Очікуваний результат: Статус змінено на «Перекладаємо». Після завершення перекладу статус змінено на перекладено і створено елемент історії відповідного типу</p>
ТС-4	<p>Назва: Переклад веб сторінки</p> <p>Кроки:</p> <ol style="list-style-type: none"> <li>1. Вставити посилання у поле для вводу</li> <li>2. Натиснути кнопку «Перекласти»</li> </ol> <p>Очікуваний результат: Статус змінено на «Перекладаємо». Після перекладу статус має бути «Перекладено». Також має бути створено елемент історії у якому буде відображено посилання та 4 кнопки для взаємодії з елементом історії (перехід та відкриття скороченої та повної версії перекладу)</p>

Шифр тест кейсу	Опис тест кейсу
ТС-5	<p>Назва: Перевірка кнопки зміни напрямку перекладу</p> <p>Кроки:</p> <ol style="list-style-type: none"> <li>1. Виконати ТС-1 з введеним напрямом перекладу.</li> <li>2. Натиснути на кнопку зміни напрямку перекладу</li> <li>3. Виконати пункти 1 та 2 з урахуванням вводу правильного тексту на обраній мові у поле вводу</li> </ol> <p>Очікуваний результат: Текст має перекладатись з обраної мови</p>
ТС-6	<p>Назва: Взаємодія з елементом історії типу «переклад тексту»</p> <p>Кроки:</p> <ol style="list-style-type: none"> <li>1. Виконати ТС-1.</li> <li>2. Натиснути на елемент історії</li> </ol> <p>Очікуваний результат: У полі вводу та виводу з'являються відповідні дані та напрямок перекладу встановлено автоматично</p>
ТС-7	<p>Назва: Перевірка елементу історії перекладу веб-сторінки</p> <p>Кроки:</p> <ol style="list-style-type: none"> <li>1. Виконати кроки з ТС-2</li> <li>2. Натиснути на гіперпосилання</li> <li>3. Натиснути на кнопку «Перейти» під категорією «Версія для читання»</li> <li>4. Натиснути на кнопку «Відкрити» під категорією «Версія для читання»</li> <li>5. Натиснути на кнопку «Перейти» під категорією «Повна версія»</li> <li>6. Натиснути на кнопку «Відкрити» під категорією «Повна версія»</li> </ol> <p>Очікуваний результат: Маємо отримати комбінований результат. Результати вказані для пунктів 2-6:</p> <ol style="list-style-type: none"> <li>1. Відкрито вірну веб сторінку.</li> </ol>

Шифр тест кейсу	Опис тест кейсу
	<p>2. Відкрито менеджер файлів з відповідним файлом версії перекладу для читання.</p> <p>3. Відкрито відповідний файл з наявним перекладом.</p> <p>4. Відкрито менеджер файлів з відповідним файлом повної версії перекладу.</p> <p>5. Відкрито відповідний файл.</p>
ТС-8	<p>Назва: Перевірка елемента історії перекладу файлу</p> <p>Кроки:</p> <ol style="list-style-type: none"> <li>1. Створити txt та docx документи. Документ з розширенням docx має містити картинку та таблиці для перекладу.</li> <li>2. Виконати ТС-3</li> <li>3. Обрати кнопку «Перейти» на відповідному контейнері елемента перекладу</li> <li>4. Обрати кнопку «Відкрити» на відповідному контейнері елемента перекладу</li> </ol> <p>Очікуваний результат: Відкрито менеджер файлів з вірним файлом для перекладу. Відкрито сам файл для перекладу. Файл має бути зі збереженням стилів відступів таблиць, параграфів.</p>
ТС-9	<p>Назва: Перевірка навігації на перекладеному веб сайті</p> <p>Кроки:</p> <ol style="list-style-type: none"> <li>1. Виконати ТС-7 до кроку 5.</li> <li>2. Перевірити навігацію веб-сторінки.</li> </ol> <p>Очікуваний результат: Навігація веб-сторінки має збігатись з навігацією веб сторінки на вході.</p>
ТС-10	<p>Назва: Перевірка фільтрів історії</p> <p>Кроки:</p>

Шифр тест кейсу	Опис тест кейсу
	<ol style="list-style-type: none"> <li>1. Виконати ТС-1, ТС-2, ТС-3 кілька разів</li> <li>2. Натискати на різні елементи історії в хаотичному порядку упродовж однієї хвилини.</li> </ol> <p>Очікуваний результат: Контейнер елементів історії перекладу має показувати відповідні результати, які збігаються з налаштуваннями історії перекладу</p>

### 3.8 Висновки

Описано функціональні та нефункціональні вимоги. Розроблено структуру системи та представлено у вигляді діаграм розгортання та компонентів. Описано сценарії використання системи та розроблено матрицю відстеження між функціональними вимогами та варіантами використання.

Також розроблено ER діаграму що описує структуру бази даних системи. Також розроблено інтерфейс користувача.

Функціональні вимоги покрито тест кейсами та створено відповідну діаграму відстеження.



## 4 МАТЕМАТИЧНА МОДЕЛЬ ПЕРЕКЛАДУ ТА ЕКСПЕРИМЕНТ

### 4.1 Поняття перекладу за допомогою маркерів

Ідея статистичного машинного перекладу походить від теорії інформації. Документ перекладається відповідно до розподілу ймовірностей  $p(\mathbf{e}|\mathbf{f})$ , де  $\mathbf{e}$  - рядок на цільовій мові (наприклад, англійська) є перекладом рядка  $\mathbf{f}$  в мові оригіналу (наприклад, французькою мовою). Проблема моделювання розподілу ймовірностей  $p(\mathbf{e}|\mathbf{f})$  була змодельована багато разів багатьма способами. Одним з підходів, який добре підходить для впровадження комп'ютера, є застосування теореми Байєса, тобто  $p(\mathbf{e}|\mathbf{f})$  пропорційне  $p(\mathbf{f}|\mathbf{e}) p(\mathbf{e})$ , де модель перекладу  $p(\mathbf{f}|\mathbf{e})$  - це ймовірність того, що джерело string є перекладом цільової рядка, а модель мови  $p(\mathbf{e})$  - це ймовірність бачити ці рядки мови. Це розкладання є привабливим, оскільки він розбиває цю проблему на дві підпроблеми. Знаходження найкращого перекладу  $\tilde{\mathbf{e}}$  здійснюється шляхом вибору тієї, що дає найбільшу вірогідність:  $\tilde{\mathbf{e}} = \arg \max p(\mathbf{e}|\mathbf{f}) = \arg \max p(\mathbf{f}|\mathbf{e})p(\mathbf{e})$ , при  $\mathbf{e} \in \mathbf{e}^*$ .

Для ретельної реалізації цього потрібно провести вичерпний пошук, пройшовши всі рядки  $\mathbf{e}^*$  рідною мовою. Ефективно виконувати пошук, це робота декодера машинного перекладу, який використовує іноземні рядки, евристику та інші методи обмеження простору пошуку з одночасним збереження допустимої якості.

Оскільки системи перекладів не можуть зберігати всі рідні рядки мови та їх переклади, документ, як правило, перекладається речення за реченням, але навіть цього недостатньо. Мовні моделі, як правило, апроксимуються моделями n-gram, і аналогічні підходи застосовуються до моделей перекладу, однак існує додаткова складність через те що довжина речень та порядок слів можуть відрізнятись.

Для того щоб навчити статистичну модель треба достатньо великий паралельний корпус текстів. Інакше для задовільної якості перекладу переклад має здійснюватися через опорну мову, наприклад англійську.

З ускладненням моделі статистичного розбору (наприклад під час використання моделей генерації правил деревовидного статистичного розбору) збільшуються і

вимоги до обчислювальної техніки для того щоб навчати мову та значно збільшується час перекладу.

Уявімо що мозок вчиться перекладати з української на англійську просто порівнянням речень без того, щоб давати певні дані. Але на пояснення усіх правил для розуміння мозком як перекладати пішло б багато часу. В такому мозку треба буде пам'ятати безліч правил та зберігати усі дані для перекладу.

Ідея маркерів полягає у тому щоб категоризувати дані які отримує система для навчання та дозволити будувати цільовий переклад не за допомогою вибору найчастіше вживаного слова або словосполучення, але вираховувати різні варіанти перекладу одного й того ж слова або словосполучення з урахуванням статистичних даних про маркери. Перелічимо маркери які можуть бути використані для навчання статистичної моделі:

- частина мови сусідніх слів;
- відмінок сусідніх іменників;
- час сусідніх дієслів;
- вид сусідніх займенників;
- слово та словосполучення на початку речення чи в кінці;
- кількість слів у реченні;
- слово або словосполучення стоїть у головному чи підрядному реченні.

Додавання даних маркерів у статистичну модель дозволить:

- розширити дані для тренування моделей які відстають по якості але не витрачають багато ресурсів з точки зору бізнесу;
- скоротити витрати ресурсів для складних моделей за рахунок категоризації знань та ліквідації повного перебору варіантів.

Переваги даної моделі полягають у наступному:

- Дана модель є ефективнішою ніж система на базі правил. Система на базі правил має бути написана для кожної мовної пари. Тобто кількість систем на базі правил які мають бути написані  $n(n-1)$ . Маркерна система потребує написання ідентифікації маркерів у речення тільки один раз для кожної мови.

- Система маркерів може бути як інтегрована у існуючу модель перекладу для тренування з нуля, так і використання у вигляді оболонки, щоб за рахунок існуючих статистичних таблиць перекладу обробити паралельні корпуси текстів ще раз та «довчити систему».
- Маркерна база даних може бути також використана для побудови якіснішого варіанту системи перекладу на базі нейронної мережі. Дані про маркери будуть давати більше знань про контекст речення, дасть можливість нейронній мережі якісніше налаштовуватись та використовувати менше ресурсів у порівняння з іншими моделями. Ще раз можна наголосити на тому що з точки зору бізнесу для кожної мови у цьому випадку все-одно треба буде написати лише одну маркерну систему.

#### 4.2 Приклад лінгвістичної бази даних

Описана база даних зібрана на кафедрі АУТС за підтримки факультету лінгвістики. Діаграма опису бази даних наведена у додатку Ж.

Таблиця анімасу описує істоту або неістоту для німецької або української мови. Під індексом 1 маємо Lebewesen (істота). Під індексом 2 маємо nicht Lebewesen (неістота). Під індексом 3 та 4 маємо істоту та неістоту відповідно для української мови. Об'єкти представлені у таблиці 4.1.

Таблиця 4.1 – Об'єкти таблиці анімасу

id	Name	language_id
1	Lebewesen	1
2	nicht Lebewesen	1
3	Істота	2
4	Неістота	2

Таблиця gens описує рід для німецької або української мови. Індeksi 1,2,3 відповідають чоловічому жіночому та середньому роду в німецькій мові відповідно. Об'єкти представлені у таблиці 4.2. Індeksi 4,5,6 відповідають чоловічому жіночому та середньому роду в українській мові відповідно.

Таблиця 4.2 – Об'єкти таблиці gens

id	Name	language_id
1	Maskulinum	1
2	Femininum	1
3	Neutrum	1
5	Чоловічий	2
6	Жіночий	2
7	Середній	2

У таблиці languages описано мови. Під індексом 1 Німецька. Під індексом 2 Українська. Мови зустрічаються як ідентифікатор у вигляді колонки таблиці під назвою language\_id.

Таблиця links описує взаємозв'язок слова в українській мові слову в німецькій мові. Колонка word\_id вказує на індекс слова у німецькій мові. Колонка translation\_id вказує на індекс слова в українській мові.

Таблиця numbers описує однину і множину в німецькій мові під індексами 1 та 2 відповідно та однину і множину в українській мові під індексами 3 та 4 відповідно. Об'єкти представлені у таблиці 4.3.

Таблиця 4.3 – Об'єкти таблиці knowledges

id	Name	language_id
1	Singular	1
2	Plural	1
3	Однина	2

id	Name	language_id
4	Множина	2

Таблиця partsOfSpeech описує частини мови. Об'єкти таблиці представлені у таблиці 4.4.

Таблиця 4.4 – Об'єкти partsOfSpeech

id	name	language_id
1	Substantiv	1
2	Adjektiv	1
3	Прикметник	2
4	Іменник	2
5	Verben	1
6	Дієслово	2
7	Adverb	1
8	Прислівник	2
10	Дієприкметник	2
11	Числівник	2
12	Numerale	1
13	Noun	3
14	Präposition	1
15	Прийменник	2
16	Konjunktion	1
17	Сполучник	2

Таблиця translations описує слово в німецькій мові. Колонка id вказує на індекс слова. Колонка name містить у собі текстову репрезентацію слова. Колонка knowledge\_id вказує на одиницю або множину. Колонка PartOfSpeech\_id вказує на

частину мови з таблиці `partsOfSpeech`. Колонка `PosStruct_id` вказує на індекс у таблиці `Pos_Ge_noun`.

Таблиця `translations` описує слово в українській мові. Колонка `id` вказує на індекс слова. Колонка `name` містить у собі текстову репрезентацію слова. Колонка `knowledge_id` вказує на одинину або множину.

Таблиця `PoS_GE_noun` описує приналежність іменника німецької мови до роду числа та істоти або неістоти. `Gen_id` колонка описує приналежність до роду. Колонка `Number_id` вказує на приналежність до числа. Колонка `Animacy_id` вказує приналежність до істоти або не істоти.

Таблиця `wordForms` описує варіанти іменників у відмінках. А таблиця `wordFormsTableStruct` описує відмінки. Відмінки наведені у таблиці 4.5.

Таблиця 4.5 – Відмінки таблиці `wordFormsTableStruct`

id	name	language_id
1	Nominativ	1
2	Genitiv	1
3	Dativ	1
4	Akkusativ	1

### 4.3 Виконання експерименту

Для підтвердження гіпотези повністю або частково доцільно провести експеримент.

#### 4.3.1 Збір паралельного корпусу

Для початку маємо зібрати паралельний корпус речень. Для цього обрано сайт `ted.com` на якому зібрані доповіді різних людей.

Порядок збору речень для паралельного словника являє собою важку задачу. Речення у кожній доповіді з німецької на українську збігаються не завжди. Іноді маємо поєднання декількох речень в одне при перекладі з мови оригіналу (англійської), тому було запропоновано наступний порядок проведення експерименту:

1. Для всіх статей відбираємо тільки ті які містять переклад з англійської на німецьку та з англійської на українську (англійська – завжди мова оригіналу доповіді).
2. Для кожної статті завантажуюмо за допомогою скрипту html файли перекладу доповіді.
3. Витягуємо абзаци з класу контейнера для тексту перекладу які знаходяться під хештегом <p> у відповідному контейнері.
4. Якщо кількість абзаців для української та німецької збігаються, це може означати, що речення на початку абзацу будуть з високою вірогідністю збігатись.
5. Збираємо з початку абзаців тільки ті речення, що збігаються по ширині не менше ніж на 85% та довжиною не більше ніж 50 символів.

За результатами маємо 3168 зібраних речень для навчання моделі.

#### 4.3.2 Опис створеної моделі

За основу візьмемо найпростішу статистичну модель IBM-1. Ця модель використовує ЕМ-алгоритм (англ. Expectation-maximization (EM) algorithm) — алгоритм, що використовується в математичній статистиці для знаходження оцінок максимальної схожості параметрів ймовірних моделей, у випадку, коли модель залежить від деяких прихованих змінних.

На виході IBM моделі маємо вірогідності кожного німецького слова бути перекладеним у українське слово. Приклад таблиці на виході моделі наведено у таблиці 4.6.

На базі цих даних маємо створити іншу таблицю вірогідності на базі обраного маркеру – частина мови слова зліва. Дані про це беруться з таблиці words описаній у підрозділі 4.2. У випадку, якщо слово не знайдено йому виставляється індекс 0. У випадку, якщо слово знаходиться на початку речення індекс буде становити -1.

Таблиця 4.6 – Приклад таблиці ймовірностей моделі IBM-1

Німецьке слово	Український переклад	Ймовірність
	легко	0.2421
	легкий	0.1242
	простий	0.0421
	...	...
	щось	0.3221
	будь-який	0.1242
	всякий	0.0121
	...	...

Процес утворення таблиці з використанням маркерів:

1. Вчимо модель на базі зібраного корпусу текстів.
2. Відсіюємо невірогідні переклади для збільшення ефективності та зрозумілості подальшого алгоритму.
3. Для кожного речення з паралельного корпусу текстів виконуємо наступне. Для кожного німецького слова рахуємо його маркер. Якщо речення української мови містить можливі переклади на прикладі таблиці 4.6, то для німецького слова записуємо не  $p(e|f)$  де  $e$  це українське слово а  $f$  – німецьке слово. а записуємо вірогідність  $p(e(M)|f)$ , де  $M$  це числове представлення маркеру. В даному випадку піднімаємо вірогідність  $e|f(M)$  на одиницю.



В результаті роботи алгоритму отримуємо 2 таблиці ймовірностей. Перша таблиця це скорочена таблиця 4.6 тільки з найімовірнішими перекладами. Приклад другої таблиці представлено у таблиці 4.7.

Таблиця 4.7 – Приклад статистичної таблиці з урахуванням маркерів

Німецьке слово	Варіант перекладу	Індекс маркеру	Ймовірність маркеру $p(e(M) f)$
		14	44
		0	21
		-1	3
...			

Вибір перекладу за допомогою моделі IBM-1 здійснюється послібно за формулою  $\tilde{e} = \arg \max p(e|f)$ .

Переклад  $\tilde{e}$  для моделі, адаптованої під маркери буде обрано наступним чином  $\tilde{e} = \arg \max (p(e|f) * p(e(M)|f))$ . Це означає що при перекладі будуть перемножені вірогідності з урахуванням частоти стрічання маркерів. Наприклад якщо для слова *leicht* в контексті частини мови зліва під індексом 14 для слова «легко» буде ймовірність маркеру (в даному випадку частота стрічання маркеру у контексті) 10 та ймовірність слова за моделлю IBM-1 0.2, а для варіанту перекладу «легкий» ймовірність маркеру 44, а за моделлю IBM-1 0.1, то буде обрано варіант перекладу «легкий», що відрізняється від перекладу обраного за допомогою натренованої моделі IBM-1. Для тестування результатів обрана метрика BLEU якості збігу вихідного речення та речення на виході.

#### 4.3.3 Результати тестування

За допомогою двох моделей перекладено 3168 речень паралельного корпусу та якість перекладу виміряна за допомогою метрики BLEU. Для моделі IBM-1 якість

перекладу склала 0.3180, а для статистичної моделі на базі маркеру 0.3672, що дало приріст показника BLEU на 15,4%.

Можна припустити, що дана модель спрацювала тільки через те, що ми дали системі нові дані для перекладу, тому вона впоралася краще, ніж та, у якої було менше даних. Для підтвердження або спростування було вирішено провести експеримент повторно але на відміну від попереднього випадку частину речень використовувати не для тренування а для перевірки метрик. Тобто кожна система буде перекладати незнайомі речення. Результати наведені у таблиці 4.8. Бачимо, що у будь-якому випадку модель на базі маркеру давала кращі результати.

Таблиця 4.8 – Залежність результатів від відсотку речень не для тренування

Відсоток речень корпусу, який використано для тренування	Параметр BLEU для моделі IBM-1	Параметр BLEU для моделі на базі обраного маркеру
100%	0.3180	0.3672
95%	0.2736	0.3280
90%	0.2687	0.3174
66%	0.2735	0.3069
33%	0.2834	0.3045

#### 4.4 Висновки

Наведено математичну модель перекладу на базі маркерів та проведено експеримент з використанням IBM моделі Python бібліотеки nltk. За результатами виконання експерименту бачимо, що навіть реалізація невеликої кількості маркерів помітно збільшує якість перекладу за метрикою BLEU.

Хоча модель перекладу розкрита не повністю, описана модель є важливим кроком для майбутніх досліджень.

## 5 РОЗРОБКА СТАРТАП-ПРОЕКТУ

### 5.1 Опис ідеї стартап-проекту

Ідея стартап-проекту описана у таблиці 5.1.

Таблиця 5.1 – Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
	Німецько-український переклад тексту	Швидкий переклад між німецькою та українською мовою без потреби мати інтернет.
	Переклад введеної веб-сторінки	Користувач має можливість виконати переклад сайту. Для цього треба всього лиш ввести текст посилання в поле де користувач також має змогу ввести текст для перекладу. Система має розпізнати гіперпосилання та автоматично згенерувати переклад сторінки у 2 форматах. Перший формат являє собою повну версію перекладу, що означає,

Зміст ідеї	Напрямки застосування	Вигоди для користувача
		<p>що кожен текст та слово, що є на сторінці перекладається і на вихід система має видати файл з розширенням .html.</p> <p>Другий формат являє собою варіант генерації файлу на виході, оптимізованого для читання, тобто на виході система має видавати .txt файл без виняткового форматування.</p>
	<p>Переклад файлів з розширенням .txt або .docx</p>	<p>Користувач має змогу обрати файл для перекладу з розширенням .txt або .docx, та отримати на виході файл з таким самим розширенням. У випадку звичайного текстового файлу в ньому перекладаються речення. У випадку вибору перекладу .docx</p>

Зміст ідеї	Напрямки застосування	Вигоди для користувача
		файлу система видає на виході такий самий файл зі збереженням форматування в якому перекладені всі параграфи та таблиці.
	Робота з інтерактивною історією	Користувач має можливість переглянути історію своїх перекладів по трьом категоріям: текст, веб-сторінки, файли. Користувач має змогу видаляти елементи історії, переходити на перекладену веб-сторінку, переходити до повної або скороченої версії перекладу, робити повтор перекладу тексту, відкривати або переходити до файлів перекладу, генерованих з вихідним файлом того самого формату. Користувач може

Зміст ідеї	Напрямки застосування	Вигоди для користувача
		сортувати історію за типом перекладу: текстовий переклад, переклад веб-сторінки, переклад файлу.

Аналіз сильних, слабких та нейтральних сторін проекту наведено у таблиці 5.2.

Таблиця 5.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проекту

		(потенційні) товари/концепції конкурентів						
		Мій проект	Google Translate	Prompt Professional 365	Microsoft Translator			
1	Вартість експлуатації	Безкоштовно з можливістю підписки на платний функціонал	Безкоштовно	2800 грн.	Безкоштовно	-	+	-
2	Кросплатформенність	-	+	+	+	+	-	-

		(потенційні) товари/концепції конкурентів						
		Мій проект	Google Translate	Prompt Professional 365	Microsoft Translator			
3	Якісний німецько-український переклад	+	—	—	—	-	-	+
4	Робота в глобальній мережі	-	+	+	+	-	+	-
5	Робота без доступу в мережу інтернет	+	—	-	-	-	-	+
6	Стійкість до відмов	Висока	Висока	Висока	Висока	-	+	-
7	Наявність системи резервування	-	+	+	+	+	-	-
8	Ступінь розширюваності	Висока	Середня	Низька	Низька	-	-	+
9	Переклад введеного тексту	+	+	+	+	-	+	-

		(потенційні) товари/концепції конкурентів						
		Мій проект	Google Translate	Prompt Professional 365	Microsoft Translator			
10	Переклад веб сторінки	+	+	+	-	-	+	-
11	Генерація .html файлу перекладеної веб-сторінки, та її автоматичне збереження	+	-	-	-	-	-	+
12	Переклад документів між німецькою та українською мовами та зберігання формату файлу на виході	+	-	-	-	-	-	+



## 5.2 Технологічний аудит проекту

Аналіз складових технологічного процесу наведено у таблиці 5.3.

Таблиця 5.3 – Технологічна здійсненність ідеї проекту

№ п\п	Ідея проекту	Технології реалізації	Наявність технологій	Доступність технологій
1		Kivy	Наявні	Доступні
2		Pyforms	Наявні	Доступні
3		PyGObject (a.k.a. PyGi)	Наявні	Доступні
4		PyQt	Наявні	Доступні
5		PyGUI	Наявні	Доступні
6		libavg	Наявні	Доступні
7		wxPython	Наявні	Доступні
Обрана технологія реалізації користувацького інтерфейсу: PyQt				
8		Бібліотека nltk	Наявні	Доступні
9		Бібліотека Moses	Наявні	Доступні
10		OpenNMT	Наявні	Доступні
Обрана технологія реалізації ідеї проекту: Спершу за допомогою nltk				
11		Miscrosoft SQL Server	Наявні	Доступні
12		ORACLE	Наявні	Доступні
13		PostreSQL	Наявні	Доступні
14		MongoDb	Наявні	Доступні
		SQLite	Наявні	Доступні
Обрана технологія реалізації ідеї проекту: SQLite				
15		Python	Наявні	Доступні
16		Java	Наявні	Доступні

17	Мова виконання	C#	Наявні	Доступні
18	свого модулю	ReactJs	Наявні	Доступні
15	перекладу	C++	Наявні	Доступні
Обрана технологія реалізації ідеї проекту: Python				

Аналіз складових технологічного процесу наведено у таблиці 5.3.

Таблиця 5.3 – Технологічна здійсненність ідеї проекту

№ п\п	Ідея проекту	Технології реалізації	Наявність технологій	Доступність технологій
1		Kivy	Наявні	Доступні
2		Pyforms	Наявні	Доступні
3		PyGObject (a.k.a. PyGi)	Наявні	Доступні
4		PyQt	Наявні	Доступні
5		PyGUI	Наявні	Доступні
6		libavg	Наявні	Доступні
7		wxPython	Наявні	Доступні
8		Kivy	Наявні	Доступні
Обрана технологія реалізації ідеї проекту: SOAP				
9		Бібліотека nltk	Наявні	Доступні
10		Бібліотека Moses	Наявні	Доступні
11		OpenNMT	Наявні	Доступні
Обрана технологія реалізації ідеї проекту: За допомогою токенів				
12		Miscrosoft SQL Server	Наявні	Доступні
13		ORACLE	Наявні	Доступні
14		PostreSQL	Наявні	Доступні
15		MongoDb	Наявні	Доступні

№ п\п	Ідея проекту	Технології реалізації	Наявність технологій	Доступність технологій
16		SQLite	Наявні	Доступні
Обрана технологія реалізації ідеї проекту: універсальний ORM EntityFramework (для варіантів 7, 8, 9)				
17		Python	Наявні	Доступні
18		Java	Наявні	Доступні
19		C#	Наявні	Доступні
20		ReactJs	Наявні	Доступні
21		C++	Наявні	Доступні
Обрана технологія реалізації ідеї проекту: Python				

В цілому, програму було вирішено писати за допомогою мови Python, тому що більшість бібліотек перекладу написані мовою Python, та з урахуванням майбутнього, бо якщо робити свій модуль роботи перекладача, то Python забезпечує велику швидкість розробки та сумісності. Для баз даних була обрана SQLite, для UI інтерфейсу обрано PyQt, для бібліотек базового перекладу nltk та OpenNmt.

### 5.3 Аналіз ринкових можливостей запуску проекту

В цілому, ринок готовий до сприйняття продукту, і продукт має свою нішу на ринку. Це підтверджується попереднім оглядом ринку, наведеному у таблиці 5.4.

Таблиця 5.4 – Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, одиниць	6
2	Загальний обсяг продаж, грн/ум.од	800 грн/ум.од
3	Динаміка ринку (якісна оцінка)	Зростає

№ п/п	Показники стану ринку (найменування)	Характеристика
4	Наявність обмежень для входу (вказати характер обмежень)	Неоднозначний характер
5	Специфічні вимоги до стандартизації та сертифікації	Потрібен патент, який захищає систему перекладу
6	Середня норма рентабельності в галузі (або по ринку), %	35%

Потенційні характеристики клієнтів стартап-проекту описані у таблиці 5.5. Це можуть бути як приватні, так і державні установи.

Таблиця 5.5 – Характеристика потенційних клієнтів стартап-проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	Переклад тексту	Вивчаючі німецьку мову у якості учнів, студентів, людей які хочуть пов'язати	Вирішальним фактором при виборі продукту є можливість його спробувати, бо продукти, які	1. Простота у користуванні 2. Висока швидкість

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
		себе з Німеччиною	можливо спробувати тільки замовивши демо-версію значно програють у кількості охочих людей, які взагалі почнуть ознайомлення з продуктом	
2	Переклад сайтів	Можлива аудиторія з пункту один та організації, які працюють з іноземними клієнтами з Німеччини	Такий сегмент користувачів цікавість надійність користування сервісом, інтуїтивність користувацького інтерфейсу якість	1. Простота у користуванні. 2. Надійність програми 3. Висока швидкість 4. Висока якість перекладу

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
			перекладу та якість виконання заявлених функцій	
3	Переклад документів	Аудиторії з пунктів 1 та 2 можуть зацікавитись продуктом.	Окремі користувачі можуть бути зацікавленими в тому, щоб у випадку завантаження документу для перекладу він не змінював свого вигляду.	Єдність стилю вихідного файлу та файлу на виході
4	Робота з історією	Менеджери, люди відповідальні за великий обсяг	Цільовим клієнтам має бути необхідно швидко рухатись по	Інтелектуальність інтерфейсу

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
		інформації, договорів, збір інформації	історії перекладів та можливість змінювати її. Для таких користувачів важлива наявність можливості швидко переходити до перекладених файлів або веб- сторінок	

Проте, український ринок в цілому і ринок інформаційних послуг в цілому є нестабільним і необхідним є врахування низки загрозливих факторів. Фактори ризику і загрози на ринку описані у таблиці 5.6.

Таблиця 5.6 – Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Непередбачувані зміни законодавства	Зміна вимог до організації з боку органів державної влади (наприклад, рівень оподаткування)	Організація юридичної особи в країні зі сприятливим інвестиційним кліматом
2	Непередбачуваний курс гривні	Ставка ІТ-працівників на українському ринку історично прив'язана до курсу іноземних валют	Фіксація заробітної плати в гривнях, співпраця з молодими спеціалістами, фіксація вартості оплатної версії програмного забезпечення у валюті.
3	Недостача стартових капіталовкладень	Може не вистачити інвестицій для реалізації мінімально життєздатного продукту або продовження	Пошук нових інвесторів



№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
		спроб вийти на ринок	
4	Недовіра користувачів	Користувачі не довіряють новому гравцю на ринку	Допомогти зрозуміти користувача допоможе зміна маркетингової стратегії та постійні опитування користувачів як з їх точки зору вони бачать продукт, що в ньому не вистачає, та що вони хотіли б бачити в майбутньому.

Проте, наразі ринок сформував також низку можливостей, реалізація яких дозволить отримати прибуток і завоювати ринок. Фактори можливостей наведені в таблиці 5.7.

Таблиця 5.7 – Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Отримання необхідних інвестицій	Сформований початковий капітал, необхідний для реалізації мінімально життєздатного продукту	Розробка мінімально життєздатного продукту
2	Великий спектр можливостей застосування перекладу	Переклад використовується у багатьох сферах діяльності (переклад картинок, переклад у реальному часі, переклад розмови у реальному часі) на різних платформах.	Розширення функціоналу, мовних пар для перекладу та платформ.
3	Успішна маркетингова та інформаційна політика	Маркетингова стратегія з прикладями успішного впровадження системи, та з висвітленням ідеології компанії.	Збільшення кількості споживачів при тій самій вартості розробки та підтримки системи.

Далі було розглянуто конкуренцію на ринку. Результати ступеневого аналізу подані у таблиці 5.8.

Таблиця 5.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
Олігополія	Незначна кількість конкурентів Велика ринкова сила Схожість використовуваних технологій	Успішна маркетингова стратегія
Галузевий	Загроза появи нових конкурентів Ринкова влада споживачів Висока потреба у товарі	Інформування ринку щодо якості використовуваної новаторської технології Пропозиція гнучких цін
Внутрішньогалузева	Діяльність в одній галузі економіки Надання сервісів одного типу	Зменшення вартості сервісу Впровадження рентабельних технологій Примноження каналів розподілу
Товарно-видова	Надання різних сервісів одного виду	Маркетингова політика

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
Цінова	Використання цін для покращення економічних умов збуту	Зменшення вартості сервісу Використання нових каналів користувачів
Марочна	Пропозиція схожого сервісу Спільна цільова аудиторія	Інформування ринку щодо якості використовуваної новаторської технології Примноження каналів розподілу

Більш детальний аналіз конкуренції на ринку було виконано за моделлю 5 сил М. Портера, результати наведені у таблиці 5.9.

Таблиця 5.9 – Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товаризамінники
Складові аналізу	Prompt, Google Translator, Microsoft Translator, Smartling	Капіталовкладення, існуючі напруження в кожній з поєднаних компонент	Відсутні	- Змінні витрати: Виробничі непрямі дегресивні - Системи інформації: реклама та	Копіювання функціонал та системи перекладу,

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
				<p>директ-маркетинг,</p> <p>- Рівень чутливості до цін: орієнтовані на цінність продукту</p> <p>- Продуктова диференціація: якість, спосіб отримання сервісу, швидкість обслуговування</p> <p>Методи контролю якості: ручне та автоматичне тестування, профілювання, прототипування, рев'ю коду, аудит</p>	Монополізація дистрибуторів, Демпінг

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
Висновки	Інтенсивність конкуренції висока. Кожен перекладач хоче зайняти свою нішу та тримати свого користувача, в той же час кожен продукт намагається розширити функціонал для того, щоб захопити інші цільові аудиторії.	Можливості входу на ринок забезпечить мінімізація цін, швидкість та простота надавання послуги споживачам, швидке реагування на вимоги користувачів. Потенційних конкурентів знайдено не було.	Постачальники відсутні, тому не диктують умов	Клієнти диктують умови гнучкості цінової політики, високої і довгострокової якості послуг, зручності користування	Пропонування вигідних умов безпосереднім користувачам, виділення ліній підтримки

Незважаючи на високу конкуренцію на ринку і наявних на ньому гравців, пропонована система має низку переваг, що описані у таблиці 5.10.

Таблиця 5.10 – Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
1	Унікальність сервісу	Висока інтегрованість конкретно німецько-українського перекладу. Наразі не існує сервісів, що концентруються на даній мовній парі.
2	Наявність моделі бізнес для бізнесу	Бізнес модель ґрунтується на співпраці із іншими підприємствами які працюють з клієнтами з Німеччини та використовують у своїй діяльності німецьку мову
3	Цінова політика	Отримання прибутку здійснюється за рахунок продажу щомісячних підписок. Звичайний типовий користувач має можливість ознайомитись безкоштовно з сервісом, а по завершенні строку безкоштовної підписки основна частина функціоналу залишається працювати, а преміум функціонал перестає працювати. При спробі взаємодіяти з преміум функціоналом буде вимагатись підписки.
4	Додаткові послуги	Можливе розширення функціональності, розширення мовних пар.

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
5	Відсутність потреби у хмарі	Більшість сервісів для перекладу розгортаються на хмарі. Це означає що користувач підключається до інтернету, надсилає дані на сервер. Сервер обробляє дані та повертає певну відповідь. В даному випадку продукт буде розгортати систему перекладу локально, що дозволить значно зеконотити на встановленні, розгортанні, масштабуванні та підтримуванні віддаленого серверу

За визначеними факторами конкурентоспроможності, наведеними у таблиці 5.10, було виконано порівняльний аналіз сильних та слабких сторін стартап-проекту з Google Translate. Результати аналізу наведені у таблиці 5.11.

Таблиця 5.11 – Порівняльний аналіз сильних та слабких сторін проекту

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з Google Translate						
			-3	-2	-1	0	+1	+2	+3
1	Унікальність сервісу	5					+		
2	Наявність моделі бізнес для бізнесу	7					+		
3	Цінова політика	15							+
4	Додаткові послуги	5					+		
5	Відсутність потреби у хмарі	16							+



Фінальним етапом аналізу можливостей на ринку з точки зору впровадження проекту є оформлення SWOT-аналізу на базі даних, отриманих раніше. SWOT-аналіз полягає у побудові матриці, що включає:

- сильні сторони проекту (Strength);
- слабкі сторони проекту (Weak);
- загрози (Troubles);
- можливості (Opportunities).

Результати проведеного SWOT-аналізу наведені у таблиці 5.12.

Таблиця 5.12 – SWOT- аналіз стартап-проекту

<p>Сильні сторони:</p> <p>Якість надання послуги</p> <p>Сучасна система монетизації</p> <p>Концентрація на німецько-українській мовній парі</p>	<p>Слабкі сторони:</p> <p>Нестача стартових капіталовкладень</p> <p>Сильна монополізація ринку</p> <p>Інертність ринку</p>
<p>Можливості:</p> <p>Інвестиції</p> <p>Реалізація бізнес-моделі</p> <p>Висока зацікавленість цільової аудиторії</p>	<p>Загрози:</p> <p>Крадіжка системи перекладу</p> <p>Вибір споживачів на користь перевірених рішень</p>

На основі SWOT-аналізу було розроблено альтернативи ринкової поведінки для виведення стартап-проекту на ринок та розраховано орієнтовний оптимальний час їх ринкової реалізації.

Визначені альтернативи були проаналізовані з точки зору термінів та ймовірності отримання ресурсів. Результати аналізу наведені у таблиці 5.13.

Таблиця 5.13 – Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Маркетингова кампанія для приваблювання користувачів	Дуже ймовірне	4 місяці
2	Безкоштовне впровадження підписки терміном на один рік для перших 100 чоловік	Дуже ймовірне	1 місяць
3	Безкоштовне впровадження системи у низці підприємств терміном у один місяць підписки	Дуже ймовірне	1 місяць
4	Присутність на великих заходах у сфері ІТ та інших у якості маркетингового партнера	Ймовірне	6 місяців
Обрана альтернатива: маркетингова кампанія для приваблювання користувачів			

#### 5.4 Розроблення ринкової стратегії проекту

Розроблення ринкової стратегії у першу чергу передбачає визначення стратегії завоювання ринку: опис цільових груп потенційних споживачів. Виконаний опис наведено у таблиці 5.14.

Таблиця 5.14 – Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Людина, яка використовує німецьку мову у повсякденні	Середня	60%		Низькі бар'єри входу
2	Компанії з німецькомовними клієнтами	Середня	30%		Високі бар'єри входу
Які цільові групи обрано: Обидві. Маркетингова стратегія буде розроблятися окремо для людей, які використовують німецьку мову у повсякденні, та для компаній з німецькомовними клієнтами					

За результатами аналізу потенційних груп споживачів (сегментів) була обидві цільові групи для яких буде впроваджена система. Стратегії охоплення ринку, наведені у таблиці 5.15.

Таблиця 5.15 – Визначення базової стратегії розвитку

№ п/ п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
1	Цільова аудиторія це	Концентрован ий маркетинг	Здатність протистояти прямим конкурентам	Стратегія спеціалізації

	людина, яка використовує німецьку мову у повсякденні	на кожній з платформ Os X, Windows відповідно на платформах додатків Microsoft Store та App Store	Низькі витрати	
2	Цільова аудиторія це компанії з німецькомовними клієнтами	Концентрований маркетинг холодного контакту	Тісна співпраця	Стратегія спеціалізації

Обраною базовою стратегією розвитку є стратегія спеціалізації, яка передбачає концентрацію на потребах одного цільового сегменту, без прагнення охопити увесь ринок. Така відмінність базується на відсутності аналогічних рішень з точки зору концентрації на німецько-українській мовній парі.

Базові стратегії конкурентної поведінки, наведені у таблиці 5.16.

Таблиця 5.16 – Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
1	Немає однозначної відповіді	Забирати існуючих у конкурентів	Модуль перекладу на базі статистичної моделі	Стратегія заняття конкурентної ніші.

Важливим моментом при виході на ринок є впізнаваність торгової марки, чому сприяє формування комплексу асоціацій, за якими користувачі можуть ідентифікувати продукт та закріпити емоційне враження за продуктом.

Для формування ринкової позиції та емоційних вражень, асоціацій, за якими споживачі будуть впізнавати проект, було визначено стратегію позиціонування, наведену у таблиці 5.17.

Таблиця 5.17 – Визначення стратегії позиціонування

№ п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
1	Простота у користуванні. Інтуїтивний інтерфейс. Достатня якість перекладу. Наявність історії перекладу. Швидкість перекладу має бути задовільна. Переклад не має залежати від мережі інтернет.	Стратегія диференціації	Формування регулярного попиту Збільшення разового використання послуги Виявлення нових груп споживачів Нові напрями застосування існуючої послуги	Німецько- український переклад. Зручний переклад веб сторінок. Зручний переклад файлів.

### 5.5 Розроблення маркетингової програми стартап-проекту

Першим кроком розроблення маркетингової стратегії стартап-проекту є формування маркетингової концепції товару, який отримає споживач. У таблиці 5.18 підсумовані результати попереднього аналізу конкурентоспроможності товару.

Таблиця 5.18 – Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Концентрація на німецько-українському перекладі	Переклад текстів, веб сторінок та документів між німецькою та українською мовами один із найкращих.	Якість перекладу Більш природній переклад
2	Автоматичне збереження HTML сторінок на виході	Швидкий доступ до перекладеного веб-сайту у вигляді файлу.	Автоматичне Збереження HTML сторінки
3	Переклад веб сторінки у файл, зручний для читання без інтерактивності	Можливість перетворити вихідну веб-сторінку у перекладений файл у вигляді версії тільки для читання	Зручність для читання перекладеного веб-сайту

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
4	Автоматичне зберігання формату та стилю вхідного файлу для перекладу	Користувач має змогу завантажити файл для перекладу у .txt або .docx форматі і має можливість отримати перекладений файл у такому ж вигляді зі збереженням форматування, формату і стилів.	Зручність перекладу .txt або .docx файлів
5	Інтерактивна історія перекладів	Користувач при перекладі здатен забувати деякі дані або використовувати їх потім, тому для нього доцільно зробити історію перекладів, яку він може переглядати, сортувати,	Наявність зручної та інтерактивної історії перекладів

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
		редагувати та взаємодіяти з елементами.	

Далі за уточненими характеристиками продукту була розроблена трирівнева маркетингова модель товару, яка наведена у таблиці 5.19.

Таблиця 5.19 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Автоматизована система німецько-українського перекладу		
	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. Дизайн 2. Зручність 3. Інтерактивність 4. Коректність алгоритмів та їх реалізації		
	Якість: коректність роботи ПЗ, відсутність нештатних ситуацій		
	Назва: Ge Uk Translator		
	До продажу: встановлення та наладка системи, навчання користувачів		
	Після продажу: реалізація додаткових функцій, розгляд та відновлення при нештатних ситуаціях		



Рівні товару	Сутність та складові
За рахунок чого потенційний товар буде захищено від копіювання: обфускація вихідних кодів та патентування системи перекладу	

Далі були оцінені принципи ціноутворення на ринку, за результатами чого були визначені межі встановлення ціни, наведені у таблиці 5.20.

Таблиця 5.20 – Визначення меж встановлення ціни

№ п/п	Рівень цін на товари- замінники	Рівень цін на товари- аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1	—	1000-5000 грн	20000 грн	50 - 600 грн

Розрахована оптимальна система збуту наведена у таблиці 5.21.

Таблиця 5.21 – Оптимальна система збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
	Закупівля здійснюється через системи маркетплейсів App Store та Microsoft Store	Просування сторінки продукту серед потенційних користувачів	Канал одного рівня	Маркетплейс

Останньою складовою маркетингової програми є розроблення концепції маркетингових комунікацій, що спирається на попередньо обрану основу для позиціонування, визначену специфікою поведінки клієнтів. Результати розробки концепції маркетингових комунікацій наведені у таблиці 5.22.

Таблиця 5.22 – Концепція маркетингових комунікацій

№ п/п	Специфік а поведінки цільових клієнтів	Канали комуніка цій, якими користую ться цільові клієнти	Ключові позиції, обрані для позиціонува ння	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Ведення діяльності і в межах державних структур	Торгові площі для програмного забезпечення від Microsoft та Apple	Доступність та об'єктивність інформації про фірму і товар Унікальність послуги Концентрація аніст на німецько- українському у перекладу	Формування у цільовій аудиторії знань про появу нового сервісу Інформування користувачів про властивості та переваги сервісу Пояснення цільовій аудиторії принципу дії послуги	Рационаліс тична стратегія реклами

№ п/п	Специфік а поведінки цільових клієнтів	Канали комуніка цій, якими користую ться цільові клієнти	Ключові позиції, обрані для позиціонува ння	Завдання рекламного повідомлення	Концепція рекламного звернення
				Інформування користувачів про нові способи використання відомого сервісу Інформування про можливість безкоштовно користуватись обмеженим функціоналом Інформування про інтелектуальніст ь сервісу	

## 5.6 Висновки

В результаті проведеної розробки стартап-проекту було сформовано список очікувань користувачів до автоматизованої системи німецько-українського

перекладу, було проведено технологічний аудит та аналіз ринкових можливостей запуску проекту, розроблено ринкову стратегію та маркетингова програма.

В цілому, є великі можливості для ринкової комерціалізації проекту. Попит на такого роду системи присутній, динаміка ринку позитивна – ринок зростає. Рентабельність роботи на ринку достатня для того, щоб займатись даним продуктом, проте ускладнена високою конкуренцією.

Є можливість впровадження системи як серед користувачів які використовують німецьку мову у повсякденні так і для організацій з іноземними клієнтами з Німеччини. Незважаючи на високу конкуренцію і монополізацію ринку, проект є конкурентоспроможним. Проект може розвиватися поетапно, поступово збираючи відгуки про продукт від користувачів. Наявні механізми розширення продукту як в глибину так і в ширину – розширення мовних пар та розширення функціоналу. Відсутність хмарних обчислень є великою перевагою.

## ВИСНОВКИ

В процесі роботи над магістерською дисертацією було розглянуто та проаналізовано існуючі системи автоматизованого перекладу. В цілому існує багато рішень для мовного перекладу з багатьма методами перекладу, але водночас можна знайти нішу для свого продукту, якщо він доведе свою конкурентоспроможність. На основі аналізу систем перекладу було розроблено автоматизовану систему німецько-українського перекладу з використанням покращеної статистичної моделі за допомогою використання маркерів. До переваг системи перед конкурентами можна віднести:

- Система концентрується на німецько-українському перекладі.
- Система розгортається у середовищі Python, що дозволяє написати єдиний код для двох систем: OS X та Windows одного боку та легко інтегрувати системи перекладу з іншого.
- Система розгортає підсистему перекладу локально, що з точки зору бізнесу вигідно для економії ресурсів на створення серверу перекладу.
- Інтерфейс користувача зроблений для максимальної легкості навігації та з присутністю історії перекладів, що дозволяє користувачеві підвищити власну ефективність.

В рамках мети описано математично модель перекладу за допомогою маркерів та проведено експеримент на реалізованій моделі з адаптуванням методу статистичного машинного перекладу. Результати експерименту позитивні, що підтверджує право моделі на існування, хоча і дана тема потребує подальшого вивчення та подальших досліджень. Головними перевагами моделі перекладу є її простота та відсутність нескінченного набору даних які треба зібрати за допомогою лінгвістів.

Результати роботи можуть бути використані для створення нових моделей перекладу та адаптації існуючих. Наведемо переваги та сфери можливого використання та подальших досліджень:

- Дана модель є ефективнішою ніж система на базі правил. Система на базі правил має бути написана для кожної мовної пари, а маркерна система тільки один раз для кожної мови.
- Система маркерів може бути як інтегрована у існуючу статистичну систему перекладу для тренування з нуля, так і використанням у вигляді побудови додаткових таблиць статистики, які приведені у розділі 4.3.2, для покращення якості перекладу.
- Теоретично маркерна база даних може бути також використана для побудови якіснішого варіанту системи перекладу на базі нейронної мережі та збільшення її швидкості навчання (при відповідній перебудові структури нейронної мережі для приймання даних про маркери).

## ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation [Електронний ресурс] / [Y. Wu, S. Mike, Z. Chen та ін.] – Режим доступу до ресурсу: <https://arxiv.org/abs/1609.08144>.
2. Quoc V. A Neural Network for Machine Translation, at Production Scale [Електронний ресурс] / V. Quoc, S. Mike – Режим доступу до ресурсу: <https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html>.
3. Sommerlad J. Google Translate: How does the search giant's multilingual interpreter actually work? [Електронний ресурс] / Joe Sommerlad – Режим доступу до ресурсу: <https://www.independent.co.uk/life-style/gadgets-and-tech/news/google-translate-how-work-foreign-languages-interpreter-app-search-engine-a8406131.html>.
4. Translate Written Words [Електронний ресурс] – Режим доступу до ресурсу: [https://support.google.com/translate/answer/6142478?hl=en&ref\\_topic=7011659](https://support.google.com/translate/answer/6142478?hl=en&ref_topic=7011659).
5. Translate text messages, webpages, or documents [Електронний ресурс] – Режим доступу до ресурсу: [https://support.google.com/translate/answer/2534559?hl=en&ref\\_topic=7011659](https://support.google.com/translate/answer/2534559?hl=en&ref_topic=7011659).
6. Translate by Speech [Електронний ресурс] – Режим доступу до ресурсу: [https://support.google.com/translate/answer/6142468?hl=en&ref\\_topic=7011659](https://support.google.com/translate/answer/6142468?hl=en&ref_topic=7011659).
7. About [Електронний ресурс] – Режим доступу до ресурсу: <https://translate.google.com/intl/en/about/>.
8. Translate images [Електронний ресурс] – Режим доступу до ресурсу: [https://support.google.com/translate/answer/6142483?hl=en&ref\\_topic=7011659](https://support.google.com/translate/answer/6142483?hl=en&ref_topic=7011659).
9. Translate with Handwriting or Virtual Keyboard [Електронний ресурс] – Режим доступу до ресурсу: [https://support.google.com/translate/answer/6142469?hl=en&ref\\_topic=7011659](https://support.google.com/translate/answer/6142469?hl=en&ref_topic=7011659).
10. Wauters R. Rant: Google Translate Toolbar In Chrome 5 Needs An 'Off' Button [Електронний ресурс] / Robin Wauters – Режим доступу до ресурсу: <https://techcrunch.com/2010/02/14/google-chrome-5-translate-toolbar/>.

11. Setalvad A. Google Translate adds 20 new languages to video text translation [Электронный ресурс] / Ariha Setalvad – Режим доступа до ресурсу: <https://www.theverge.com/2015/7/29/9061135/google-translate-update-new-languages-word-lens>.
12. Petrovan B. Google Translate just got smarter: Word Lens and instant voice translations in latest update [Электронный ресурс] / Bogdan Petrovan – Режим доступа до ресурсу: <https://www.androidauthority.com/google-translate-word-lens-voice-translations-580312/>.
13. Olanoff D. Google Translate's App Now Instantly Translates Printed Text In 27 Languages [Электронный ресурс] / Drew Olanoff – Режим доступа до ресурсу: <https://techcrunch.com/2015/07/29/google-translates-app-now-instantly-translates-printed-text-in-27-languages/>.
14. Burnette E. Google pulls the rug out from under web service API developers, nixes Google Translate and 17 others [Электронный ресурс] / E. Burnette – Режим доступа до ресурсу: <https://www.zdnet.com/article/google-pulls-the-rug-out-from-under-web-service-api-developers-nixes-google-translate-and-17-others/>.
15. Google Translate Help [Электронный ресурс] – Режим доступа до ресурсу: <https://support.google.com/translate#topic=7011659>.
16. Shankland S. Google Translate now serves 200 million people daily [Электронный ресурс] / Stephen Shankland – Режим доступа до ресурсу: <https://www.cnet.com/news/google-translate-now-serves-200-million-people-daily/>.
17. Translate text messages, webpages, or documents [Электронный ресурс] – Режим доступа до ресурсу: <https://support.google.com/translate/answer/2534559>.
18. Translate by speech [Электронный ресурс] – Режим доступа до ресурсу: <https://support.google.com/translate/answer/6142468>.
19. What is neural network based translation? [Электронный ресурс] – Режим доступа до ресурсу: <https://microsofttranslator.uservoice.com/knowledgebase/articles/1099027-what-is-neural-network-based-translation>.



20. Wendt C. Microsoft Translator Hub: Discussion of BLEU Score [Електронний ресурс] / C. Wendt, W. Lewis – Режим доступу до ресурсу: <https://www.microsoft.com/en-us/research/video/microsoft-translator-hub-discussion-of-bleu-score/?from=http%3A%2F%2Fresearch.microsoft.com%2Fapps%2Fvideo%2Fdefault.aspx%3Fid%3D193716>.
21. Microsoft Translator launching Neural Network based translations for all its speech languages [Електронний ресурс] – Режим доступу до ресурсу: <https://blogs.msdn.microsoft.com/translation/2016/11/15/microsoft-translator-launching-neural-network-based-translations-for-all-its-speech-languages/>.
22. Microsoft Translator - Customization [Електронний ресурс] – Режим доступу до ресурсу: <https://www.microsoft.com/en-us/translator/business/customization/>.
23. Microsoft Translator live feature [Електронний ресурс] – Режим доступу до ресурсу: <https://www.microsoft.com/en-us/translator/business/live/>.
24. Where Language Meets the World: Microsoft Translator Hub [Електронний ресурс] – Режим доступу до ресурсу: <https://www.youtube.com/watch?v=UnROjeEHqDE&list=PLD7HFcN7LXRd4kd2XgZjIbQ8TwTC32Zc9&index=9>.
25. Міщенко А. Л. Лінгвістика фахових мов та сучасна модель науково-технічного перекладу / А. Л. Міщенко. – Вінниця: Нова книга, 2013. – 448 с.
26. Vauquois B. The approach of GETA to automatic translation. Comparison with some other methods / Bernard Vauquois. // International Symposium on Mechanical Translation. – 1985. – С. 67.
27. Как работает нейронный машинный перевод? [Електронний ресурс] – Режим доступу до ресурсу: <https://habr.com/users/Lokalise/>.
28. Horwood E. Machine translation: past, present, future. 1986. Chichester.
29. Dorr B.J., Jordan P.W., Benoit J.W. A Survey of Current Paradigms in MT, University of Maryland, College Park, Technical Report 1998.
30. Hutchins W.J. Current and potential applications of machine translation // Hutchins Web. 2012. URL: <http://www.hutchinsweb.me.uk/Aslib-2012.pdf>

31. Statistical Post-Editing of a Rule-Based Machine Translation System [Электронный ресурс] / A. L.Lagarda, V. Alabau, F. Casacuberta, R. Silva – Режим доступа до ресурсу: <http://www.aclweb.org/anthology/N/N09/N09-2055.pdf>.
32. Bangalore S., Haffner P., Kanthak S. Statistical Machine Translation through Global Lexical Selection and Sentence Reconstruction // Proceedings of ACL. Prague. 2007.
33. Zhou S. Has AI surpassed humans at translation? Not even close! / Sharon Zhou. // "Has AI surpassed humans at translation? Not even close!". – 2018.
34. Syntax-based Statistical Machine Translation / P.Williams, R. Sennrich, M. Post, P. Koehn., 2016. – 208 с.
35. Wołk K. Real-Time Statistical Speech Translation / K. Wołk, K. Marasek. // Advances in Intelligent Systems and Computing. – 2014. – С. 107–114.
36. Sutskever I. Sequence to sequence learning with neural networks / I. Sutskever, O. Vinyals., 2014.
37. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches / C.Kyunghyun, B. van Merriënboer, D. Bahdanau, Y. Bengio., 2014.
38. Hutchins J. Machine translation: A concise history / J. Hutchins. // Computer aided translation: Theory and practice. – 2007.
39. Boretz A. AppTek Launches Hybrid Machine Translation Software [Электронный ресурс] / Adam Boretz // Speech Technology. – 2012. – Режим доступа до ресурсу: <http://www.speechtechmag.com/Articles/News/News-Feature/AppTek-Launches-Hybrid-Machine-Translation-Software-52871.aspx>
40. Carbonell J., Klein S., Miller D., Steinbaum M., Grassian T., Frey J. Context-based Machine Translation // Proceedings of 7th ACL in the Americas. Cambridge, Ms. 2006.
41. Koehn P., Hoang H., Birch A., Callison-Burch C., Federico M., Bertoldi N., Cowan B., Shen W., Moran C., Zens R., et al. Moses: Open Source Toolkit for Statistical Machine Translation // Annual Meeting of the Association for Computational Linguistics. Prague. 2007.
42. Kalchbrenner N. Recurrent Continuous Translation Models / N. Kalchbrenner, P. Blunsom. // Proceedings of the Association for Computational Linguistics. – 2013.

ДОДАТОК А  
Скан акту впровадження



АКТ ВПРОВАДЖЕННЯ

Даний акт підтверджує впровадження результатів роботи, проведених в дипломному проєкті студента «Київського Політехнічного Інституту імені Ігоря Сікорського» Кота Юрія Анатолійовича на тему "Автоматизована система двомовного перекладу за допомогою маркерів". Основні результати роботи: створена модель перекладу з використанням маркерів та німецько-український паралельний корпус впроваджені у складі інноваційної діяльності ТОВ "АЙМАКС МЕНЕДЖМЕНТ УКРАЇНА".

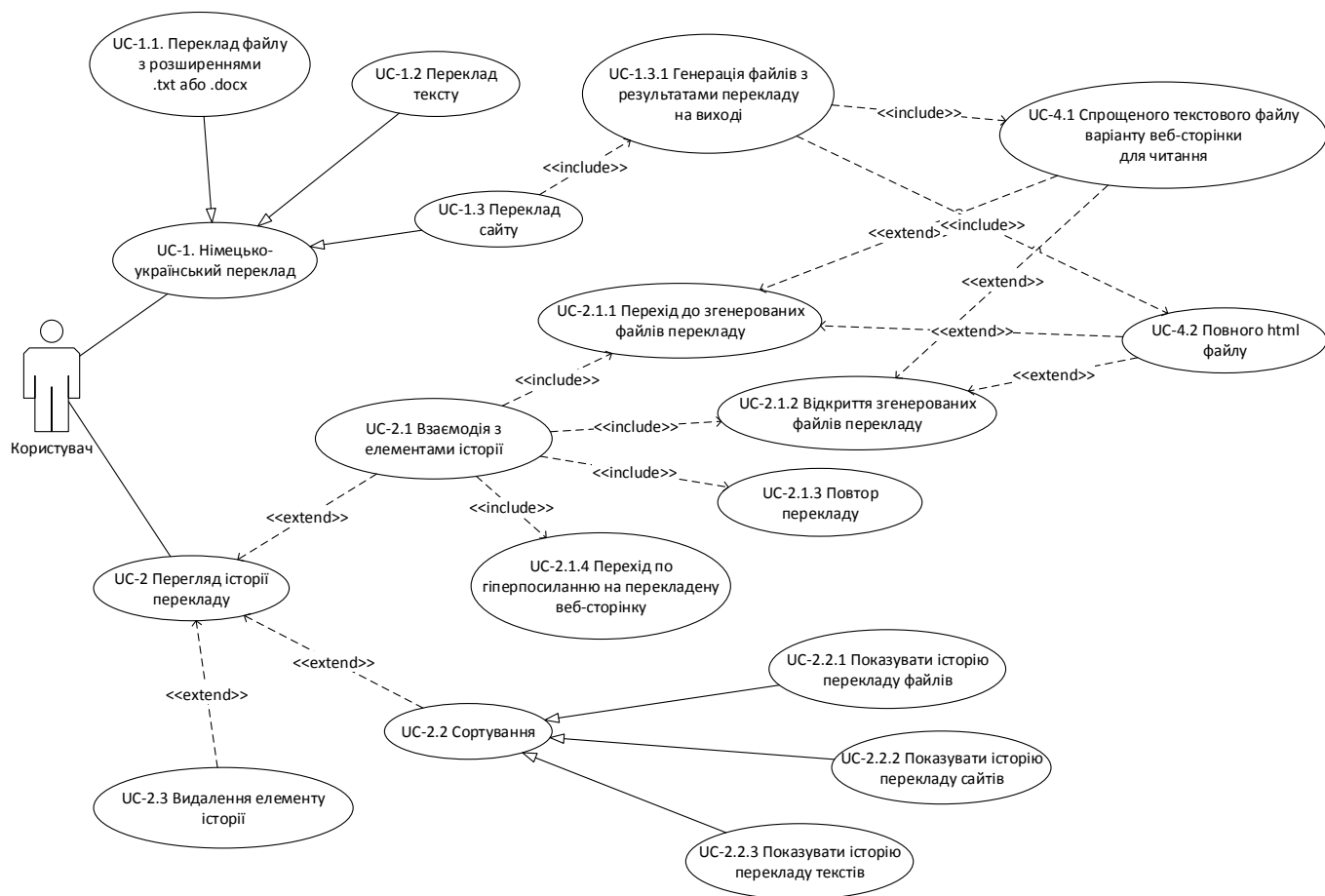
Директор ТОВ  
"АЙМАКС МЕНЕДЖМЕНТ УКРАЇНА"



Самелюк Ю. В.

## ДОДАТОК Б

### Діаграма варіантів використання



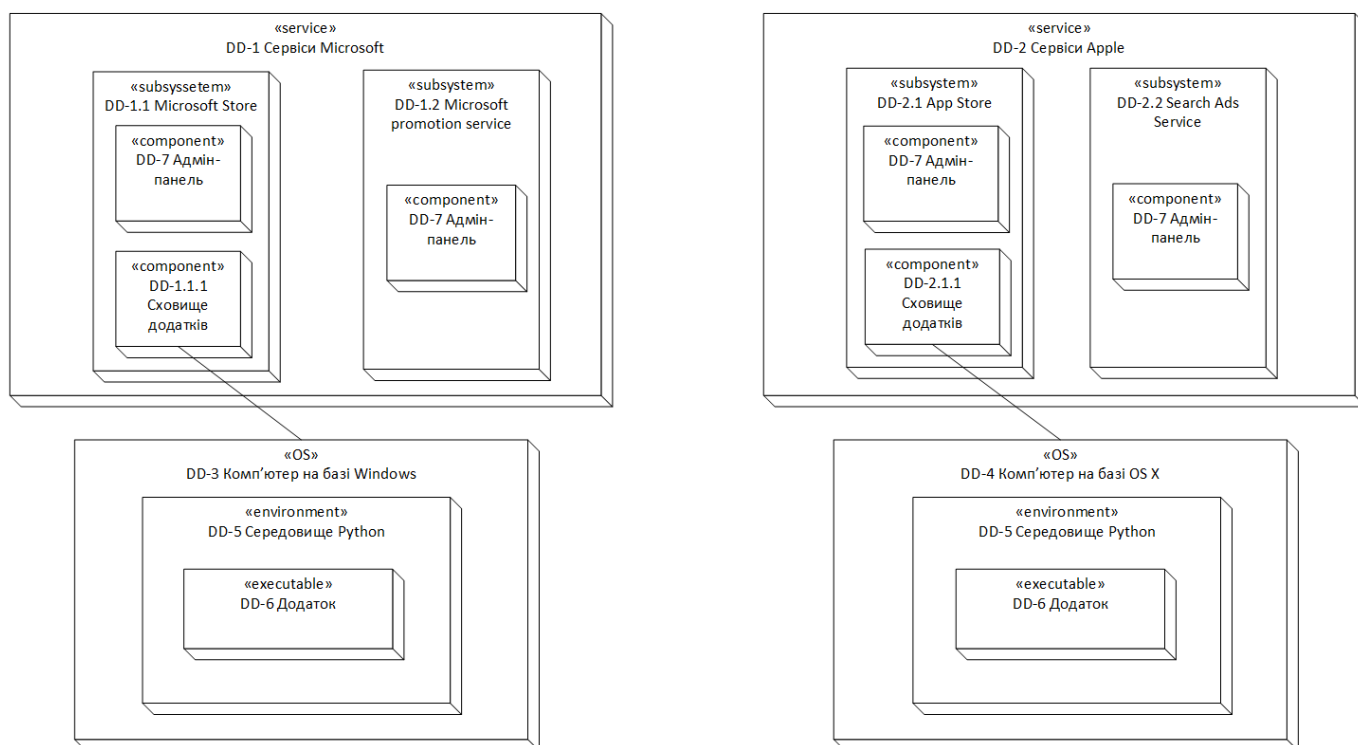
## ДОДАТОК В

### Матриця відстеження варіантів використання

[illegible]

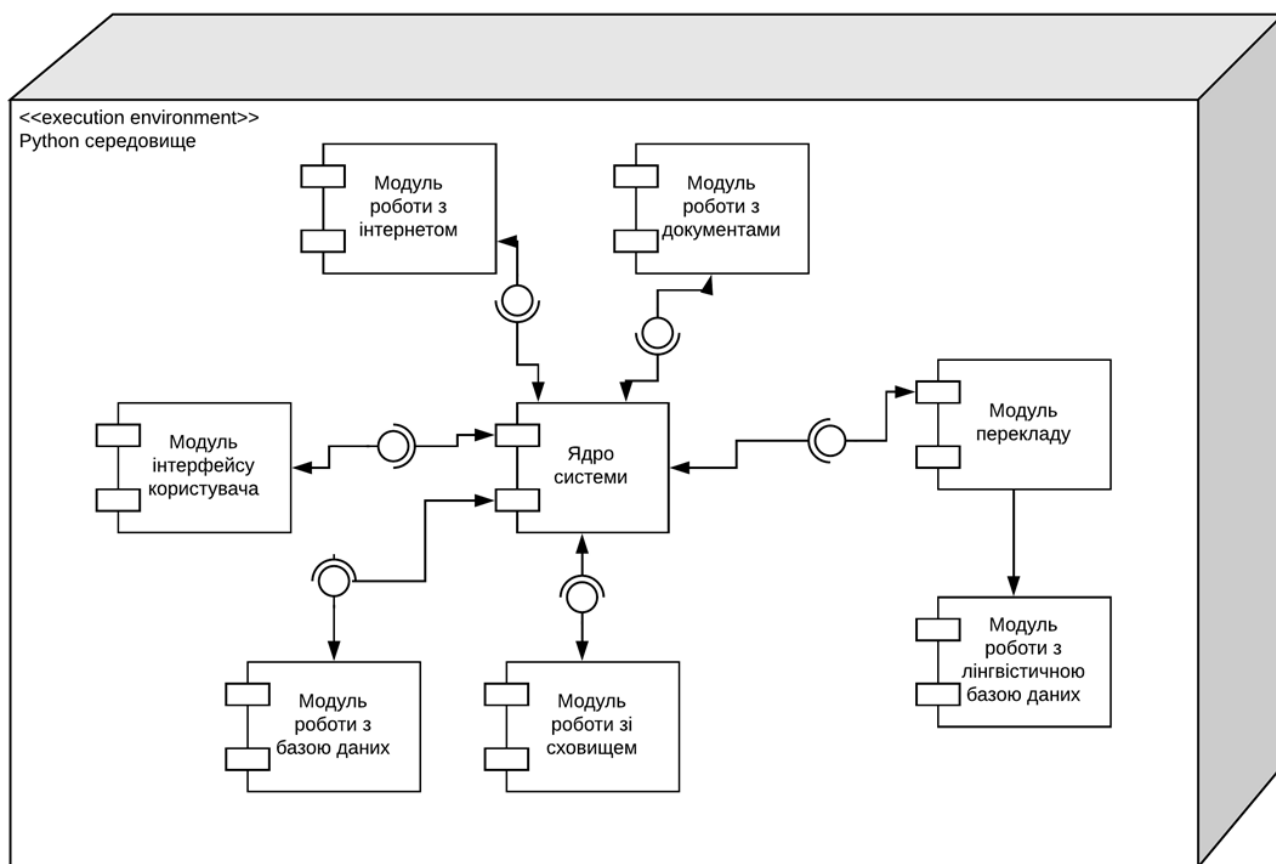
## ДОДАТОК Г

### Діаграма розгортання

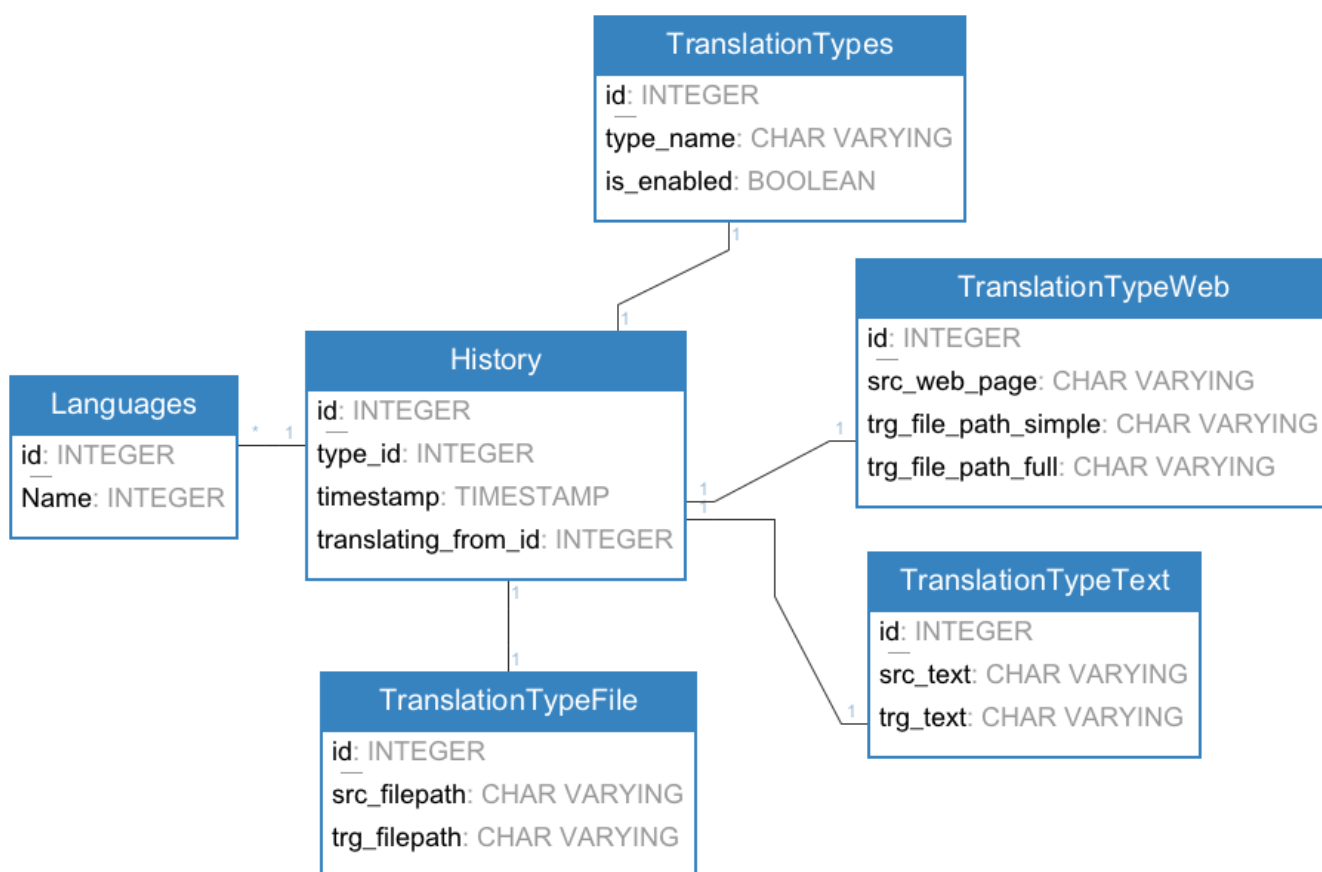


## ДОДАТОК Г

### Діаграма компонентів



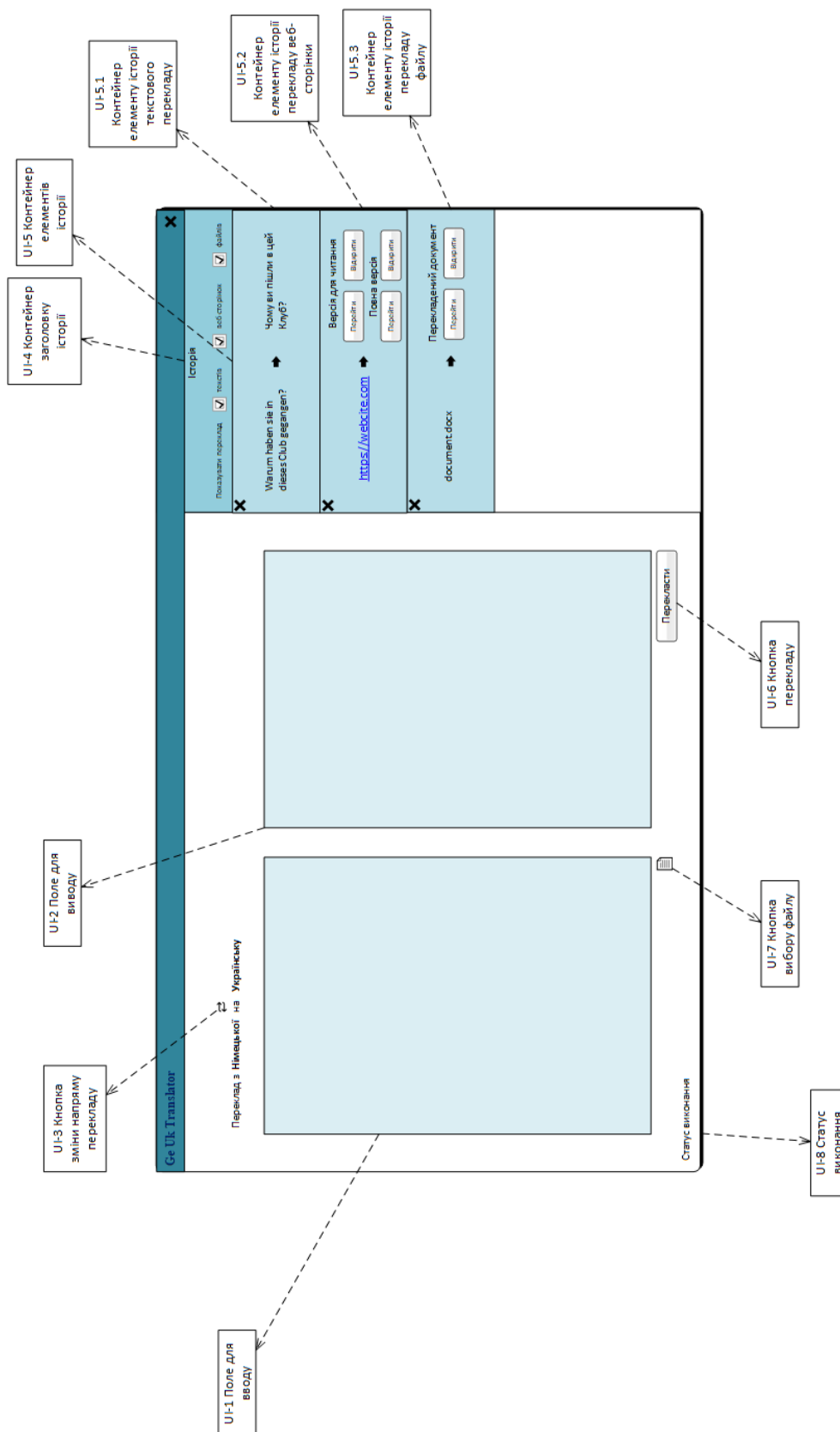
ДОДАТОК Д  
ER діаграма бази даних





## ДОДАТОК Е

### Макет інтерфейсу користувача



ДОДАТОК Є

Матриця відстеження тест-кейсів

Шифр вимоги	Кількість тест кейсів на вимогу	Шифр тест-кейсу									
		ТС-1	ТС-2	ТС-3	ТС-4	ТС-5	ТС-6	ТС-7	ТС-8	ТС-9	ТС-10
		Кількість вимог на тест кейс									
2	0	2	2	1	1	1	1	1	2	1	
FR-1	2	x					x				
FR-2	2					x				x	
FR-3	1				x						
FR-4	8	x			x			x		x	x

# ДОДАТОК Ж

## ER діаграма бази даних модулю перекладу

